Sound Mosaics

A GRAPHICAL USER INTERFACE FOR SOUND SYNTHESIS BASED ON AUDITORY-VISUAL ASSOCIATIONS

A thesis submitted to Middlesex University
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Konstantinos Giannakis

School of Computing Science

Middlesex University
United Kingdom

December 2001

Abstract

This thesis presents the design of a Graphical User Interface (GUI) for computer-based sound synthesis to support users in the externalisation of their musical ideas when interacting with the system in order to create and manipulate sound.

The approach taken consisted of three research stages. The first stage was the formulation of a novel visualisation framework to display perceptual dimensions of sound in visual terms. This framework was based on the findings of existing related studies and a series of empirical investigations of the associations between auditory and visual percepts that we performed for the first time in the area of computer-based sound synthesis. The results of our empirical investigations suggested associations between the colour dimensions of *brightness* and *saturation* with the auditory dimensions of *pitch* and *loudness* respectively, as well as associations between the multidimensional percepts of *visual texture* and *timbre*.

The second stage of the research involved the design and implementation of *Sound Mosaics*, a prototype GUI for sound synthesis based on direct manipulation of visual representations that make use of the visualisation framework developed in the first stage. We followed an *iterative design* approach that involved the design and evaluation of an initial Sound Mosaics prototype. The insights gained during this first iteration assisted us in revising various aspects of the original design and visualisation framework that led to a revised implementation of Sound Mosaics.

The final stage of this research involved an evaluation study of the revised Sound Mosaics prototype that comprised two controlled experiments. First, a comparison experiment with the widely used *frequency-domain* representations of sound indicated that visual representations created with Sound Mosaics were more *comprehensible* and *intuitive*. Comprehensibility was measured as the level of accuracy in a series of sound-image association tasks, while intuitiveness was related to subjects' response times and perceived levels of confidence. Second, we conducted a formative evaluation of Sound Mosaics, in which it was exposed to a number of users with and without musical background. Three usability factors were measured: *effectiveness*, *efficiency*, and *subjective satisfaction*. Sound Mosaics was demonstrated to perform satisfactorily in all three factors for music subjects, although non-music subjects yielded less satisfactory results that can be primarily attributed to the subjects' unfamiliarity with the task of sound synthesis.

Overall, our research has set the necessary groundwork for empirically derived and validated associations between auditory and visual dimensions that can be used in the design of cognitively useful GUIs for computer-based sound synthesis and related areas.

Acknowledgements

First and foremost, I would like to thank my supervisors, Prof. Ann Blandford and Dr. Matt Smith for providing all the time and support needed to complete my studies. It has been an incredible experience to work with you, learn from you, and be inspired by you. Thanks are also due to the research team at Middlesex University who gave me the opportunity and the resources to do this research.

I would also like to thank all the students and staff at the Sonic Arts Dept. (Middlesex University) as well as my fellow research students who participated in the experiments described in this thesis. Particular thanks go to Dr. John Dack, Andrew Deakin, Martin Robinson, and Tony Gibbs for making the Sonic Arts Dept. feel like a second home.

Special thanks to Thomas Tan and Penny Duqenoy for their support and friendship. Other people who I would like to thank for making my time at Middlesex University such a pleasant experience include: Casper Nielsen, Doreen Ng, Kok Fong Tan, Sardia Alhassan, and Dr. Serengul Smith.

My interest in sound synthesis and novel approaches to music composition would not have been the same without some inspiring discussions I had with my best friend and artist Nicolas Teloglou.

My heartfelt thanks to Nasia for her love and patience all these years. Finally, I cannot find words to express my love and gratitude to my family for their unconditional love and support throughout my life. This is dedicated to you.

Contents

For my parents

1

Introduction

"Most composers today use tools developed by themselves that are well suited for their specific goal. There is still a lack of generalized tools that would free composers from the burden of developing such tools. Therefore, they require an *expertise* that has *little* to do with composition."

[Holtzman 1994, p. 168, emphasis added]

1.1 The Research Problem

The research described in this thesis is concerned with sound synthesis by means of computers. Computers can generate sounds either for the imitation of acoustic instruments or the creation of new sounds with novel timbral properties. Almost 50 years of research in computer music laboratories world-wide has resulted in the development of a large body of diverse sound synthesis techniques, for example additive synthesis, subtractive synthesis, physical modelling, and granular synthesis (detailed discussions of these techniques can be found in Roads (1996), Dodge and Jerse (1997), and Miranda (1998)). Usually, a synthesis technique comprises a set of low-level parameters related to Digital Signal Processing (DSP) modules (*unit generators* in computer music jargon) such as oscillators, filters, amplifiers, and so on. A common characteristic of synthesis techniques is that a sound is represented as an object consisting of a large number (hundreds or thousands) of short sub-events that can be controlled by numerous time-varying parameters. Inevitably a vast amount of musical data must be defined and modified making the process of creating a sound object a very complex, non-musical and tedious task. Therefore, although it is theoretically possible to create almost any sound using one or more techniques, a fundamental question in the design of computer-based systems for sound synthesis is how to better support users in the externalisation of their musical ideas when *interacting* with the system in order to create and manipulate sound.

The most common approach to address this interaction problem has been the design of user interfaces (e.g. text-based, graphic, or other) for the control of synthesis parameters. It is useful here to place the discussion within the context of a human-computer system in which the components of the interface facilitate interactions between the user and the system as illustrated in Figure 1.1.

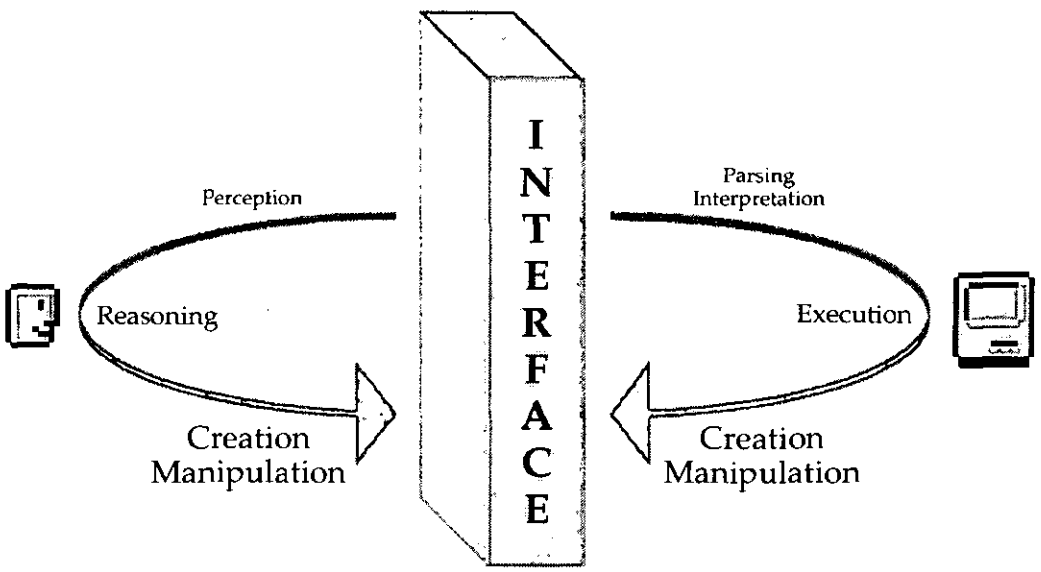


Figure 1.1: Human-Computer interaction cycle based on Narayanan and Hübscher (1998).

Figure 1.1 is based on a model suggested by Narayanan and Hübscher (1998), in which the interface component takes the form of a *visual language*. However, for the purposes of our research, the interface part takes the form of a Graphical User Interface (GUI). According to Narayanan and Hübscher, a GUI may be viewed as a visual language with an alphabet consisting of visual representations. Although the term 'visual language' has been mainly used to describe programming languages whose syntax is based on visual representations, Narayanan and Hübscher argued that "current theories of visual languages need to be extended to include a larger, more interesting class of languages" (p. 92).

The idea of using visual representations in computer-based sound synthesis is not new. For example graphical editors (e.g. for the drawing of waveforms) and on-screen interconnections of DSP objects are common features in current sound design tools reviewed in the next chapter. In this thesis, we argue that current visual representations used in the design of GUIs for sound synthesis systems focus more on the computational aspects (right half of Figure 1.1) and less on cognitive aspects (left half of Figure 1.1) of the above interaction cycle. These representations are based on low-level characteristics of sound (for example, a sound waveform depicts the variation of amplitude over time)

and cannot be used as intuitive sources for sound synthesis. In other words, it is impossible to infer the acoustic result from the representation itself. Furthermore, the proposed associations between auditory and visual dimensions, i.e. the visualisation frameworks underlying these representations, have not been empirically derived or validated. As a result of these limitations, users have shifted their focus from the high-level musical task of sound design to the low-level and cumbersome process of understanding and controlling the visualisation framework idiomatic to each representation.

1.2 The Thesis

The research described in this thesis focuses on the cognitive aspects of human-computer interaction in the context of computer-based sound synthesis. My central thesis is the following. *Visual representations of sound that take advantage of empirically derived cognitive associations between high-level dimensions of auditory and visual perception can form an adequate theoretical framework for the design of cognitively useful graphical user interfaces for computer-based sound synthesis tools.*

The approach taken has been the design and implementation of *Sound Mosaics*; a prototype GUI for sound synthesis based on direct manipulation of visual representations. We have taken an interdisciplinary approach, offering a novel visualisation framework for the associations between auditory and visual percepts that has been formulated upon the findings of existing related studies and a series of empirical investigations that we performed for the first time in the area of computer-based sound synthesis. This framework sets the necessary groundwork for the design of visual representations of sound that are based on empirically derived (as opposed to arbitrary) auditory-visual associations.

An important aspect of this thesis is an attempt to incorporate a model of auditory perception in the process of sound design. This model draws on the findings of existing studies in psychoacoustics, as discussed in more detail in Chapter 2. Since the formulation of a complete model of auditory perception is a formidable task, the scope of this thesis is confined to the auditory dimensions of *pitch, loudness,* and dimensions of *timbre* that pertain to the steady-state portions of sounds. In addition, this thesis proposes a limited set of visual dimensions — namely dimensions of *colour* and *visual texture* — for the control and manipulation of the above auditory dimensions that draws on existing studies in visual perception.

As a result of this approach, Sound Mosaics allows users to create and manipulate sound in perceptual terms, thus making the process of sound synthesis a less complex and more intuitive activity than the manipulation of low-level sound characteristics.

1.3 Motivation

The significant role of visual communication in modern computer applications is indisputable. As stated by Colin Ware in his recent book on information visualisation, "The human visual system is a pattern seeker of enormous power and subtlety. The eye and the visual cortex of the brain form a massively parallel processor that provides the highest-bandwidth channel into human cognitive centers. At higher levels of processing, perception and cognition are closely interrelated, which is the reason why the words 'understanding' and 'seeing' are synonymous" (Ware 2000, p. xviii).

In the case of music it seems that it is very natural for musicians to translate non-visual ideas into visual codes (see Walters (1997) for examples of graphic scores from J. Cage, K. Stockhausen, I. Xenakis, and others). Also, in recent years, associations between auditory and visual elements have inspired a new artistic movement under the title of *visual music* (see for example Wells (1980), Goldberg and Schrack (1986), De Witt (1987), Peacock (1988), Karinthi (1991), Whitney (1980, 1991), and Pocock-Williams (1992)). Furthermore, investigations of the phenomenon of *synaesthesia* and cross-modal associations in the field of cognitive psychology have provided further support for the analogies between different sensory experiences (see §2.1.3).

Although there has been such a cross-disciplinary interest in the investigation and application of visual metaphors for musical purposes, the quite different research methodologies followed in the above scientific and artistic domains have not facilitated interdisciplinary dissemination and co-ordination of research efforts. The proposed auditory-visual associations are primarily based on subjective judgements rather than on empirical evidence. In addition, most research effort to date has focused on the *macro-compositional* level, i.e. the arrangement of sound objects into a musical score, overlooking the design of new timbres. In fact, the dimension of timbre has been largely neglected and oversimplified. This is not surprising if we take into consideration the fact that traditional music compositional processes have focused mainly on pitch and duration, and treated timbre as a second-order attribute of sound (Wishart 1996). As a result, there is no theoretical framework for auditory-visual associations that is based on empirical studies and that can be used for intuitive sound descriptions. The lack of such a framework formed the motivation for our research efforts.

1.4 Structure of Thesis Document

This chapter provided the background, motivation and objectives of our research. In Chapter 2, we review a number of visual representations of sound as these have been used in current GUIs for sound synthesis and we discuss various auditory-visual associations that have been proposed in related research areas (vision research and cognitive psychology).

In Chapter 3, we review existing research in the area of auditory perception. One of the goals of the review is to define a model of auditory perception that can be incorporated in computer-based sound synthesis and further investigations of the associations between auditory and visual dimensions.

Chapters 4 and 5 taken together form the empirical core of our research. Chapter 4 begins with a review of colour perception followed by the design and results of an empirical investigation of the associations between colour dimensions and the auditory dimensions of pitch and loudness. In a similar manner, Chapter 5 reviews studies of visual texture perception and presents an empirical study of the associations between dimensions of visual texture and timbre.

In Chapter 6, the results of our empirical investigations are combined to formulate the theoretical framework behind the design of Sound Mosaics. We review our visualisation framework and present the implementation details of an initial Sound Mosaics prototype.

A comparison study between visual representations of sound created with Sound Mosaics and the widely used *frequency-domain* representations is reported in the first part of Chapter 7 followed by a formative *usability* evaluation of Sound Mosaics. The results of our empirical studies formed the basis for a revised implementation of Sound Mosaics as described in the first part of Chapter 8. The results of a second comparison study and a final usability evaluation of Sound Mosaics are discussed in the remainder of Chapter 8.

Finally, Chapter 9 makes concluding remarks about the scientific contributions of our thesis and presents suggestions for further work following on from the research presented in this thesis.

2

Related Work

This chapter is divided into two main sections. In the first section we present a critical review of existing auditory-visual associations for music compositional processes. Various issues related to the visualisation of auditory information as well as weaknesses and limitations of the above studies have been identified and we reflect on them in a general discussion presented in the second section of this chapter. In more detail, the sources of these limitations are discussed together with ways that our research suggests as suitable to overcome them.

2.1 Auditory-Visual Associations for Music Compositional Processes

Music composition can be seen as the product of two distinct but complementary processes: (i) the design of individual sound objects, i.e. the *micro-compositional* level and (ii) the arrangement of sound objects into a musical score, i.e. the *macro-compositional* level. Although this distinction is not always clear (especially in the case of computer music), it is used here to outline the scope of our research. As mentioned in the introduction to this thesis, computer-based sound synthesis has paved the way for the exploration of novel sound spaces, although the exploration means are complex and require great expertise. Although the ultimate goal of computer music research is to encompass all music compositional processes, our primary goal in this thesis is to facilitate the specification and manipulation of sound objects at the micro-compositional level before these objects can be further used in other compositional levels.

There is a large number of studies and attempts to correlate auditory and visual elements and these can be classified in three main categories:

- Graphic sound synthesis.
- Colour spaces for sound.
- *Synaesthesia*, and cross-modal associations.

2.1.1 Graphic Sound Synthesis

According to Roads (1996) "*graphic sound synthesis* characterises efforts that start from a visual approach to sound specification" (p. 329). In general, graphic sound synthesis allows various sound characteristics to be specified, manipulated and modified via graphical means.

Current approaches in graphic sound synthesis are primarily based on visual representations of sound that fall into two main categories:

- Time-domain representations.
- Frequency-domain representations.

The above representations are discussed in detail in the remainder of this section.

Time-Domain Representations

Sound is the result of air disturbances produced by the excitation of some object. What we hear is the result of air pressure variation in our ears. A time-domain representation of sound depicts the variation of air pressure over time in the form of a two-dimensional graph with air pressure and time on the vertical and horizontal axes respectively (see Figure 2.1). This variation, called the waveform of a sound, can be *periodic* (i.e. it demonstrates a repetitive pattern), *noise* (i.e. no repetitive pattern is discernible) or fall in between these two extremes (quasi-periodic or quasi-noise). One repetition of a periodic waveform is called a *cycle*, and the *fundamental frequency* of the waveform is the number of cycles that occur per second. The amount of air pressure change at each point in time specifies the intensity of the sound, i.e. its amplitude.

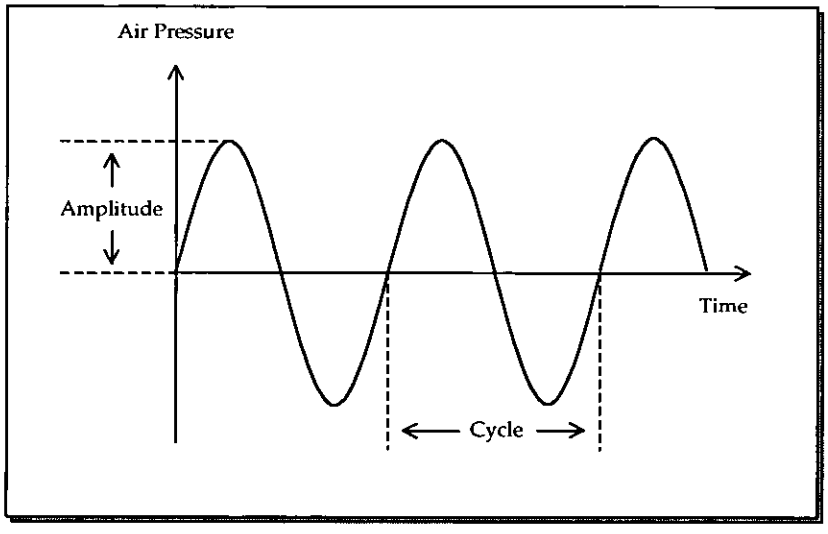


Figure 2.1: Generic form of time-domain representations.

In the context of computer-based sound synthesis, air pressure is usually plotted against phase instead of time (see Figure 2.2). In this manner, only one cycle of the waveform needs to be specified that is repeated according to a particular repetition rate. Alternatively, various waveform cycles can be specified and collated in order to form a more complex waveform as in *waveform segment techniques* (Roads 1996). One cycle of a waveform includes 360° of phase and the phase of a particular point on a waveform is

measured as an angle from some reference position, most commonly taken as the point where the waveform has a value of zero and is increasing (Dodge and Jerse 1997).

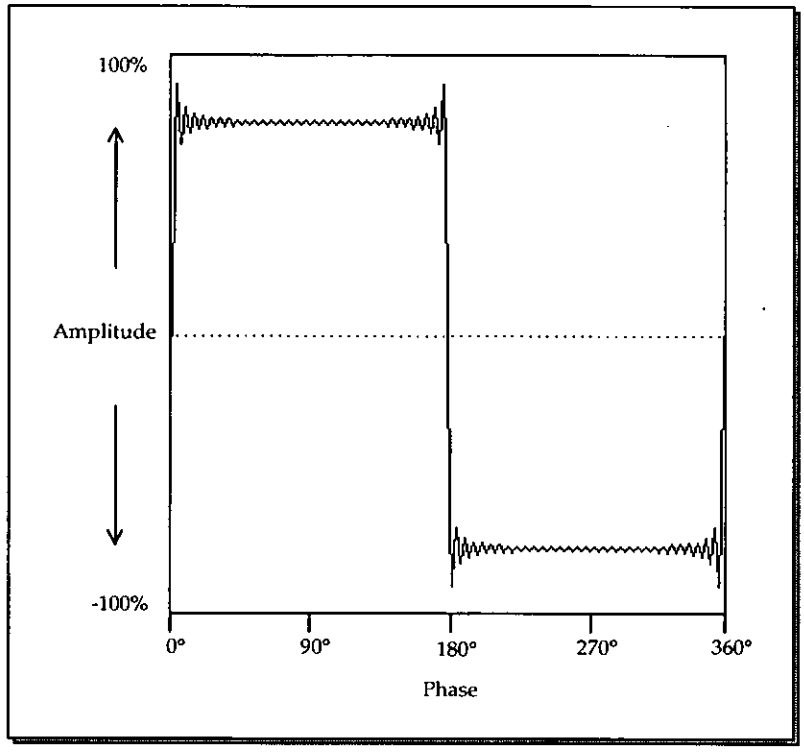


Figure 2.2: Time-domain representation with air pressure plotted against phase.

Graphical editors for the drawing and manipulation of waveforms can be found in many computer-based sound synthesis tools such as the *UPIC* system (Xenakis 1992), *Turbosynth* (Digidesign 1995), and *MetaSynth* (U & I Software 1998). Although time-domain representations can be useful in examining the pattern of amplitude variation over time, there is a lack of direct relationship between the depicted amplitude variation and the perceptual attributes of sound. Waveforms are representations of the physical dimensions of sound disregarding the ways we actually perceive sound. As a result, it requires great expertise to describe a waveform in perceptual terms or to tell how a waveform will sound. In addition, two waveforms may look different but sound identical, i.e. they have different phase relationships among their frequency components but exactly the same spectral content (Roads 1996). Finally, although users can use a variety of tools to draw a desired waveform it is very hard to design a waveform precisely by hand (Nelson 1997).

Based on the above it can be argued that time-domain representations are counterintuitive tools in computer-based sound synthesis.

Frequency-Domain Representations

A frequency-domain representation of sound is a graph showing the individual frequency components (called partials) that comprise a sound as well as their relative amplitudes. Usually, there is a sound analysis stage prior to the representation that is commonly based on the well-known *Fourier* analysis, although, other methods may also be used such as the *wavelet transform* (see Roads (1996), Pierce (1999)). According to Fourier, any complex waveform can be decomposed into a series of sinusoidal waveforms (see Figure 2.3) with different frequencies and amplitude levels. When these sinusoids are added together the resulting waveform is the initial complex one.

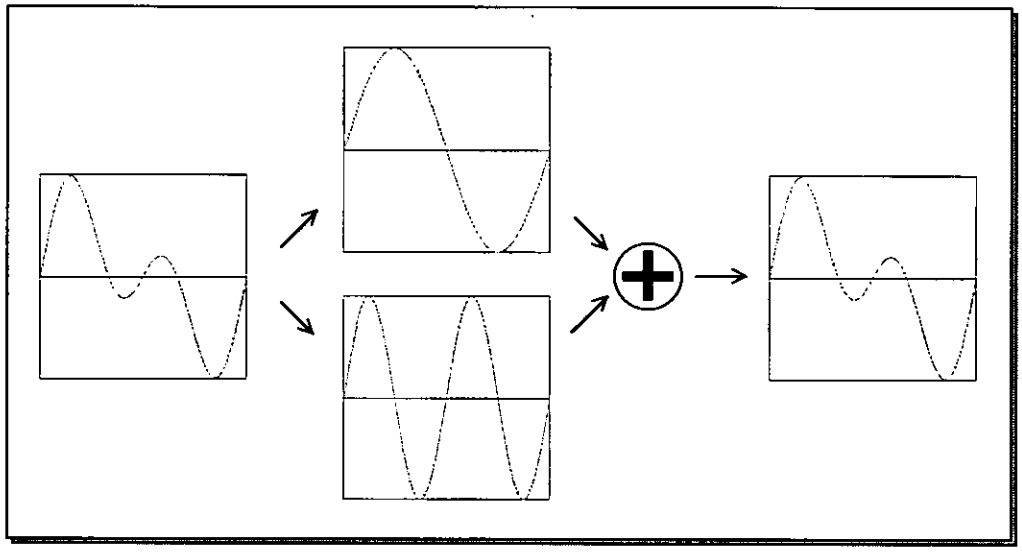


Figure 2.3: The complex waveform shown on the left can be decomposed through Fourier analysis into a series of sinusoidal components (middle) that when added together result in the initial complex waveform (right).

For periodic sounds, the frequencies of the individual sinusoids are integer multiples of the fundamental frequency (for example, with a fundamental frequency of 110 Hz, the first multiple will be 220 Hz, the second multiple 330 Hz, the third multiple 440 Hz, etc.). These integer multiples of the fundamental are also called *harmonics*. In the case of quasi-periodic and noise sounds, the frequencies of the individual sinusoids need not be integer multiples of the fundamental. An important assumption behind Fourier analysis is that the auditory signal is assumed to be periodic. Therefore, the analysis of quasi-periodic or noise sounds is problematic (Roads 1996).

Frequency-domain representations can be either *static* or *time-varying*. A static frequency-domain representation of sound is a two-dimensional image showing the frequency content of the sound at a specific point in time. Frequency and amplitude are usually

plotted on the horizontal and vertical axes respectively. Representations that have been taken at different points in time can be combined together on a single two-dimensional image thus showing all the frequency components that occurred during all the analysis stages. However, since the evolution of the frequency components over the duration of the sound is not included in the representation, these representations may be misleading. For example, the sound of a piano note will have the same visual representation as when it is played backwards because the spectral content of both sounds is the same although the two sounds are very different (Risset and Wessel 1999). Static frequency-domain representations have been used in sound synthesis tools that allow users to specify the desired frequency-amplitude pairs (see Figure 2.4).

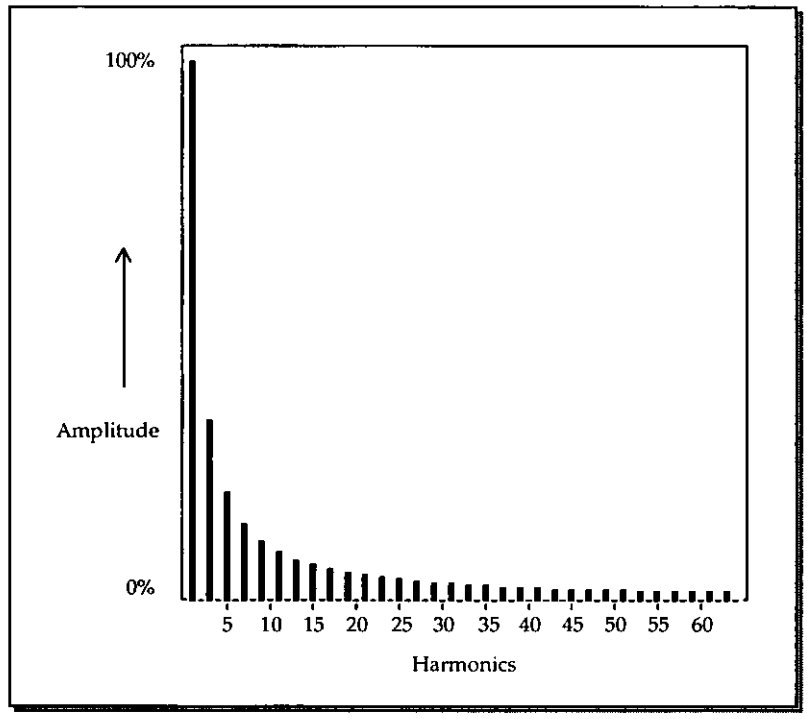


Figure 2.4: Static frequency-domain representation as used in Turbosynth (Digidesign 1995).

Time-varying frequency-domain representations of sound depict the variation of the frequency content over the duration of the sound thus giving a better picture of sound evolution than static representations. In this case, the three spatial dimensions x, y, and z are usually used to represent frequency, amplitude, and time respectively, as shown in Figure 2.5. However, these representations have been used in analysis tools and not for the purposes of sound synthesis. Another way to display a time-varying spectrum is to plot a *sonogram* (or *spectrogram*). A sonogram shows the frequency versus time content of a sound, where frequency and time are plotted on the vertical and horizontal axes respectively, and the amplitudes of the frequencies in the spectrum are plotted in terms of the darkness of the trace (see Figure 2.6).

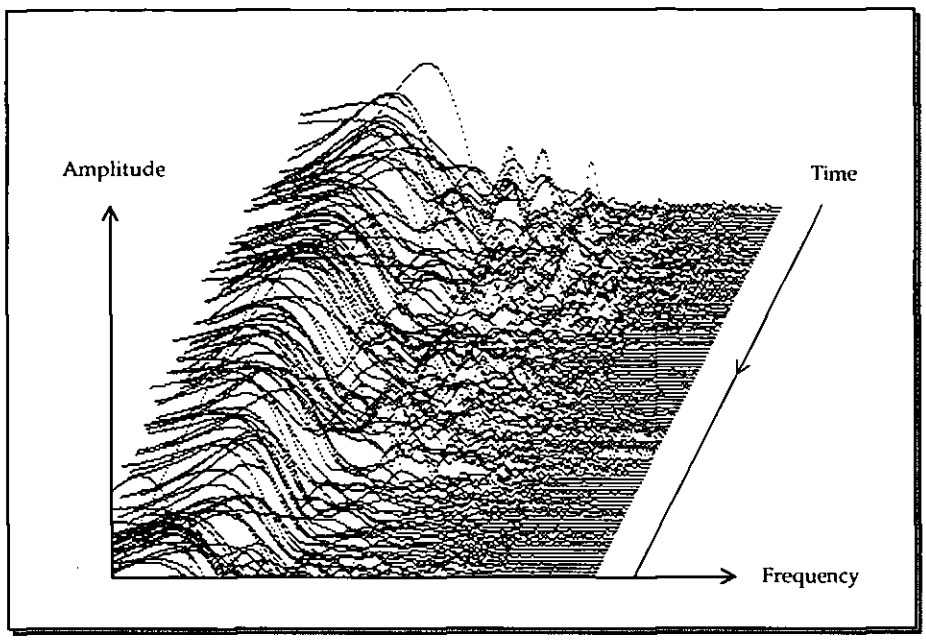


Figure 2.5: Three-dimensional frequency-domain representation. Amplitude and frequency are on the vertical and horizontal axes respectively. Time runs on the z-axis (back to front).

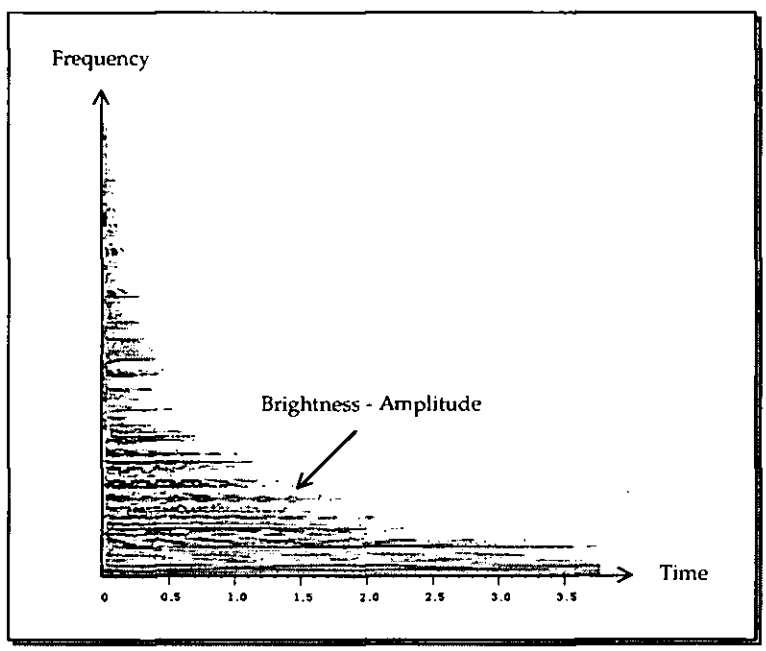


Figure 2.6: A sonogram is a moving image with frequency and time represented on the vertical and horizontal axes respectively. The brightness of the trace represents the amplitude of the frequencies (image produced with Audiosculpt (IRCAM 1995)).

Sonogram representations have been widely used in sound design tools such as those described in the remainder of this section.

Lemur (Fitz, Haken and Holloway 1995) is a sound synthesis tool based on the analysis of sampled sounds. *Lemur* performs a series of short-time Fourier analyses of a sampled sound in order to extract the frequency content of the sound, which is then reduced to the most significant partials using the *McAulay-Quartieri* algorithm as described in Fitz, Walker and Haken (1992). These partials are then plotted in a similar manner to sonograms with frequency and time plotted on the vertical and horizontal axes respectively with the brightness of the partial tracks representing their amplitudes at each point in time. Users can then directly manipulate the visual representation in order to perform various modifications such as scaling and shifting of the components. Similar approaches can be found in *SpecDraw* (Eckel 1992) and *Audiosculpt* (IRCAM 1995).

Phonogramme (Lesbros 1996) is a graphic editor that translates images to sounds. The underlying image-to-sound representation is called a *phonogram* that resembles sonogram representations of sound as described above. Lesbros (1996) experimented with different kinds of physical drawing tools and techniques (e.g. ink, pencil, watercolour, etc.) to create drawings that were scanned and used as raw materials for graphic sound synthesis. However, we believe that these techniques require deeper interpretation. For example, the purpose of drawing with watercolour is to create a certain visual effect. This means that Phonogramme lacks a higher level of image processing that is required for interpreting our drawing and producing a similar acoustic result.

MetaSynth (U & I Software 1998) is a graphic sound synthesis and music composition environment that makes it possible to synthesise sounds directly from an image coming from any source (e.g. user drawn, graphics file or derived from the analysis of an existing sound). At the level of sound design, Metasynth performs a Fast Fourier Transform (FFT) on a source sound and produces a sonogram representation that can be altered and manipulated by applying various image filters. At the macrocompositional level, a picture is scanned from left to right. Frequency and time are represented on the vertical and horizontal axes respectively. A red-yellow-green scale is used to determine the spatial position of sound while the greyscale level of pixels specifies the amplitude as in sonogram representations.

Frequency-domain representations allow us to examine various relations among the various frequency components that determine the *timbre* of the sound. As discussed in more detail in §3.1.3, timbre has been shown to depend on spectrum characteristics and therefore frequency-domain representations give at a theoretical level a better picture of various sound attributes than time-domain representations. However, this is still a task that requires great expertise.

2.1.2 Colour Spaces for Sound

In this section, we review research efforts that are inspired from colour perception studies and attempt to describe sound in terms of a small set of colour dimensions. An important concept in colour perception research is that of *colour space*. The latter is a formal method of representing the visual dimensions of colour (Jackson *et al* 1994). There are various examples of colour spaces ranging from purely physical models such as the *RGB* to more perceptually based models (e.g. *HSV, CIELUV, NCS*) and these are discussed in more detail in §4.1. Colour spaces present a number of important features that are also highly desirable in the area of sound design. For example, by arranging colours in a three-dimensional space it is easy to understand concepts such as colour complementarity, similarity, and contrast. In a similar manner, it may be possible to structure sound relations in terms of similarity, difference, and so on.

Barrass (1997) cites Padgham (1986) and Caivano (1994) as the first attempts to model sound using colour spaces. These studies are primarily based either on arbitrary auditory-visual mappings or on correspondences that may exist between the physical dimensions of sound and colour. For example, in Caivano's approach, hue is associated with pitch since both these dimensions are closely related to the dominant wavelengths in colour and sound spectra respectively. In the same manner pure (or high-saturated) colours are associated with pure (or narrow bandwidth) tones whereas low-saturated colours (those that involve wider bandwidths of wavelengths) are associated with complex tones and noise. Finally, brightness is associated with loudness (black and white represent silence and maximum loudness respectively with the greyscale representing intermediate levels of loudness). Barrass (1997) experimented with various mappings between auditory and colour dimensions and proposed a three-dimensional sound space where the auditory dimensions of timbre, brightness, and pitch are associated with colour hue, saturation, and brightness respectively. These auditory-visual associations were employed in the design of a GUI for sound selection called *SoundChooser*. It is of further interest to investigate whether the above associations can be supported by empirical studies.

Sebba (1991) has taken a different approach. In a series of experiments, she investigated the structural correspondences that may exist between colour and music elements (as opposed to direct comparison of the elements themselves). The experiment results suggest that such structural correspondences between colour and music do exist (e.g. emotional expression, hierarchical organisation, contrast).

An important limitation of the above studies is that they treated the dimension of timbre as a uni-dimensional attribute of sound. Various empirical studies (for example Bismarck (1974a), Grey (1975), Ehresman and Wessel (1978), McAdams (1999)) have shown that

timbre is a multidimensional attribute of sound and have proposed a small number of prominent dimensions for the qualitative description of timbre (see §3.1.3). Therefore, the dimension of timbre has been oversimplified in existing approaches to describe sound in terms of colour dimensions.

2.1.3 Synaesthesia and Cross-Modal Associations

One useful source of information for auditory-visual associations may be a closer investigation of the phenomenon of *synaesthesia*. Harrison and Baron-Cohen (1997) define synaesthesia as "occuring when stimulation of one sensory modality automatically triggers a perception in a second modality, in the absence of any direct stimulation to this second modality" (p. 3). Synaesthesia is a rare phenomenon with recent estimates ranging from 1/25000 to 1/1000000 adults and is more common among women than men (Dann 1998). Although associations between various sensory modalities have been reported (for example gustatory hearing, tactile hearing, coloured smell), the association between visual and sonic stimuli (i.e. coloured hearing synaesthesia) is one of the most common synaesthetic conditions and manifests itself in two different but very related phenomena:

- *Coloured music*, i.e. visual sensations produced by musical sound.
- *Coloured vowels*, i.e. visual sensations produced by the sound of vowels.

In both phenomena, synaesthetes experience different musical notes (or vowels) as different colours (for example, the note *C* may be red, *D* may be blue, the vowel *u* may be perceived as green, and so on), although these experiences are rarely in agreement among different synaesthetes. Furthermore, high-pitched stimuli evoke lighter colours while low-pitched stimuli are experienced as darker colours.

In one of the most detailed accounts of synaesthesia to date, Marks (1975) examined a large number of reported synaesthesia studies related to coloured vowels and combined the results in order to identify general characteristics and consistencies among synaesthetes. Marks used the *opponent* colour model (see Fairchild (1994), Jackson *et al* (1994), §4.1) with the opponent colour axes being: black-white, red-green, and yellow-blue. He found that the black-white axis predicted vowel pitch and that the red-green axis predicted the ratio of the first two formants in the vowel spectra (the first two formants are considered to be the most important ones for vowel discrimination). Marks also reported experiments with non-synaesthete subjects and musical tones that have shown associations between pitch and lightness as well as loudness and lightness (auditory-visual associations made by non-synaesthete subjects are generally described as *cross-modal* associations). Although, the pitch-lightness association has been also reported in recent empirical studies with non-synaesthetes (e.g. Hubbard (1996)) and it is

in agreement with earlier synaesthesia studies, Marks' overall conclusion was that it is neither pitch nor loudness that is related to lightness, but auditory *brightness*. This conclusion was based on an assumption that auditory brightness is the same as auditory *density*, a dimension that increases when both pitch and loudness increase. However, auditory brightness has been shown to be a dimension of timbre that is determined by the upper limiting frequency and the way energy is distributed over the frequency spectrum of a sound (see Bismarck (1974), Grey (1975), §3.1.3). Furthermore, a problem lies in the method behind the above-described experiments. Marks investigated only the dimension of lightness, therefore other perceptual dimensions of colour such as hue and saturation were not considered. It is not very surprising that when people are asked to relate either pitch or loudness to a dark-light scale, they will succeed in both pitch and loudness. The question that arises is what happens when there are multiple visual and auditory dimensions for the subjects to associate.

2.1.4 Summary

Although graphic sound synthesis supports the attempts for a visual metaphor to sound design, various weaknesses and limitations hamper current research in this area. Visual representations of sound such as time-domain and frequency-domain representations are based on physical approaches to sound understanding and cannot be used as intuitive conceptual metaphors for sound design. No attempt has been made to investigate the associations between visual dimensions and perceptually based characteristics of sound. Colour dimensions have been incorporated in a number of current computer music systems for graphic sound synthesis with the most common association being between brightness and loudness. Other colour dimensions such as hue and saturation have been neglected with the exception of Metasynth where a red-yellow-green hue scale is used to determine the spatial position of sound. However, none of the reviewed audio-visual mappings is based on empirical evidence. This limitation has resulted in a number of different approaches that in certain cases are very different and inconsistent (for example, there is no general agreement on the use of dark (or light) as soft or loud).

Colour dimensions have been investigated to a greater extent in the reviewed attempts to create colour spaces for sound although this area also lacks the empirical evidence to support or validate the proposed correlations between auditory and colour dimensions. In addition, prominent dimensions of sound such as timbre have been oversimplified, thus further limiting the scope of those approaches.

The correspondences between auditory and visual stimuli proposed by synaesthesia studies can be summarised as *pitch class-colour hue* and *pitch height-colour lightness*, where pitch class refers to the interval position within an octave and pitch height is a monotonic dimension representing the overall pitch level in a scale from low to high (see Shepard

(1999), §3.1.1). Studies of cross-modal associations have shown similar associations for non-synaesthete subjects and it can be argued that the associations between auditory and visual stimuli can be generalised to people without the condition of synaesthesia. However, although the above associations are empirically supported, various methodological problems have been identified (e.g. other colour dimensions have not been investigated in the reported experiments).

It becomes evident from the reviewed work in the previous sections that the visualisation of auditory information is an active area in a number of different disciplines. Although there has been such a cross-disciplinary interest in the investigation and application of visual metaphors for musical purposes, the quite distinct research methodologies incorporated in the above disciplines have not facilitated interdisciplinary attempts and co-ordination of research efforts. In general there is no theoretical framework for auditory-visual associations based on empirical studies and that can be used for intuitive sound descriptions.

2.2 Discussion

In recent years, there has been a growing interest in the use of visualisation for the design of interactive systems (see for example Shneiderman (1998), Narayanan and Hübscher (1998), Ware (2000)). The goal here is to develop visual representations that are effective in the communication of information to users. We agree with Narayanan and Hübscher when they assert that "three central issues of information representation are *what* is to be represented, *how* to represent it, and how to *associate* the representation with the represented" (p. 97).

The first issue refers to what aspects or characteristics of sound we are interested in representing and a distinction can be made here between *physical* and *perceptual* characteristics of sound. So far, we have discussed visual representations of sound such as time-domain and frequency-domain representations that function at a low level of abstraction representing physical parameters of sound. Although these representations are better than textual representations it is not guaranteed that they are suitable for the task of sound synthesis. For example, an analysis of a one-second sound file sampled at a rate of 44.1 kHz yields 44100 sample values that become very difficult to interpret when viewed in textual form. However, if these values are represented graphically on a two-dimensional plane with amplitude and time on the vertical and horizontal axes respectively, various data relationships can be immediately observed and the editing and modification of data become much easier than for textual representations. If we turn the above analysis issue into a synthesis problem, then it becomes clear that the specification of a large number of values by textual means is cumbersome, and graphical interaction is

much preferred. However, although the specification and manipulation of sound data is significantly facilitated through graphical means, the question that arises is to what extent the represented information matches the users' intentions (or mental image of the sound). This thesis argues that the auditory parameters of interest are perceptually based and to this end we have to look into studies of auditory perception in order to gain useful insights about which dimensions of sound should be available for users to control.

The second issue refers to what visual dimensions should be used for the visualisation of auditory dimensions. The visualisation framework that underlies a representation can be *arbitrary, sensory,* or a mix of arbitrary and sensory mappings (Ware 2000). As Ware points out, "the word *sensory* is used to refer to symbols and aspects of visualizations that derive their expressive power from their ability to use the perceptual processing power of the brain without learning. The word *arbitrary* is used to define aspects of representation that must be learned, having no perceptual basis" (p. 10). Although current visual representations of sound are based on sensory mappings (for example, spatial and colour dimensions) it is postulated here that the visualisation of auditory information might present a different challenge. This is related to the third issue in the development of visual representations that refers to how the representation and what is being represented are associated. Based on our review of existing approaches in the visualisation of auditory information it can be argued that sensory mappings (e.g. spatial and colour dimensions) have been used in an *ad hoc* fashion without any empirical support and validation. However, auditory information could be different from ordinary data (e.g. a series of stock market prices) in the sense that it is directly related to a different modality. As such, there might be perceptually based correspondences between different modalities, in this case hearing and vision, which are hard to break and need to be identified. Partial evidence for such auditory-visual associations comes from the reviewed investigations of synaesthesia and cross-modal associations.

In conclusion, it can be argued that the limitations of current research in this area are a function of two factors. First, the auditory dimensions that have been incorporated in the reviewed studies correspond to *low-level* characteristics of sound. Second, visual dimensions have been used in an *ad hoc* fashion without any empirical support or validation.

In order to address these limitations, our research suggests that perceptually based dimensions of sound should be incorporated into the design of computer-based sound synthesis tools and such dimensions can be identified by examining various studies of auditory perception. These dimensions can be visually represented by using various sensory percepts such as colour, shape, texture, and motion among others. The exact auditory-visual mapping should be based on carefully conducted empirical investigations.

Finally, mention should be made that the visualisation of auditory information is not the only approach in high-level interface design for computer-based sound synthesis. Notable examples are attempts to design sound synthesis systems that interact with users at a high level of abstraction based on verbal descriptions of sound (e.g. *SeaWave* by Ethington and Punch (1994), *ARTIST* by Miranda (1994). However, although we have a very rich vocabulary to describe sound, concise verbal descriptions are rarely possible (especially in the case of unique sounds, where no acoustic counterpart exists) and their meaning could be very ambiguous. Furthermore, users of these systems are either forced to use a predefined vocabulary (as in SeaWave) or in the case where the vocabulary is user-configurable (as in ARTIST) users must be able to associate their perceptual experiences with the low-level parameters of the underlying synthesis mechanism. As stated by Arnheim (1974), "Language cannot do the job directly because it is no direct avenue for sensory contact with reality; it serves only to name what we have seen or heard or thought (...) it refers to nothing but perceptual experiences. These experiences, however, must be coded by perceptual analysis before they can be named" (p. 2).

3

The Perception of Sound

In this chapter, we discuss the properties of some fundamental dimensions of auditory perception that are of interest in the area of computer-based synthesis. Our goal is to derive a model of auditory perception that will assist us to determine *what* characteristics of sound need to be visualised in order to allow users to control and manipulate sound in perceptual terms. To this end, we present a review of various studies in *psychoacoustics*, the scientific field concerned with the relationship between the physical attributes of auditory stimuli and their subjective perception by humans.

3.1 Towards a Model of Auditory Perception

3.1.1 Pitch

Pitch is our perception of frequency (the number of times that the waveform of a sound repeats itself in a second) and is defined by the American Standards Association (ASA) as "that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale" (ASA 1960). For a pure tone (a sinusoidal waveform), pitch is determined only by the repetition rate of the waveform. However, the pitch of complex sounds is related to the fundamental frequency and the harmonic structure of the sound (Plomp and Rasch 1999).

Pitch is a subjective quantity as opposed to the physical quantity of frequency. Psychological experiments that test people's judgements of tones with different frequencies have resulted in perceptual scales for pitch, such as the *mel* scale (Stevens, Volkmann and Newman 1937). However, Rasch and Plomp (1999) have criticised the *mel* scale as being ambiguous and unreliable and they suggested the physical frequency scale measured in Hertz (Hz) as a rough indication of our pitch sensation. The human auditory mechanism can detect frequencies in the range from 20 Hz - 20000 Hz, although these limits can vary with the listener's age, health, gender, and other factors (Butler 1992). In addition, our pitch sensitivity is higher in the range from 200 Hz - 2000 Hz (Dodge and Jerse 1997).

Pitch is usually treated as a one-dimensional attribute arranged in a scale from low to high. This scale comprises approximately 1400 just-noticeable differences in pitch throughout our hearing range (Olson 1967). However, there is strong evidence that pitch should be treated as a two-dimensional attribute of sound with the two dimensions being *pitch height* and *tone chroma* (or pitch class). This is primarily based on the perceptual similarity between tones whose frequencies are separated by octaves, i.e intervals in the ratio of 2:1 (or a power of 2:1). The results of various studies of musical systems used in different cultures suggest that octave equivalence is culturally universal (Deutsch 1999), (Shepard 1999). An octave interval can be subdivided into any number of sub-intervals (for example, in the Western music tradition the octave is usually divided in twelve equal

steps as in *equal temperament* tuning) and tone chroma refers to the interval position within the octave. Pitch height is a monotonic dimension representing the overall pitch level in a scale from low to high. The two-dimensional approach of pitch perception is illustrated in Figure 3.1.

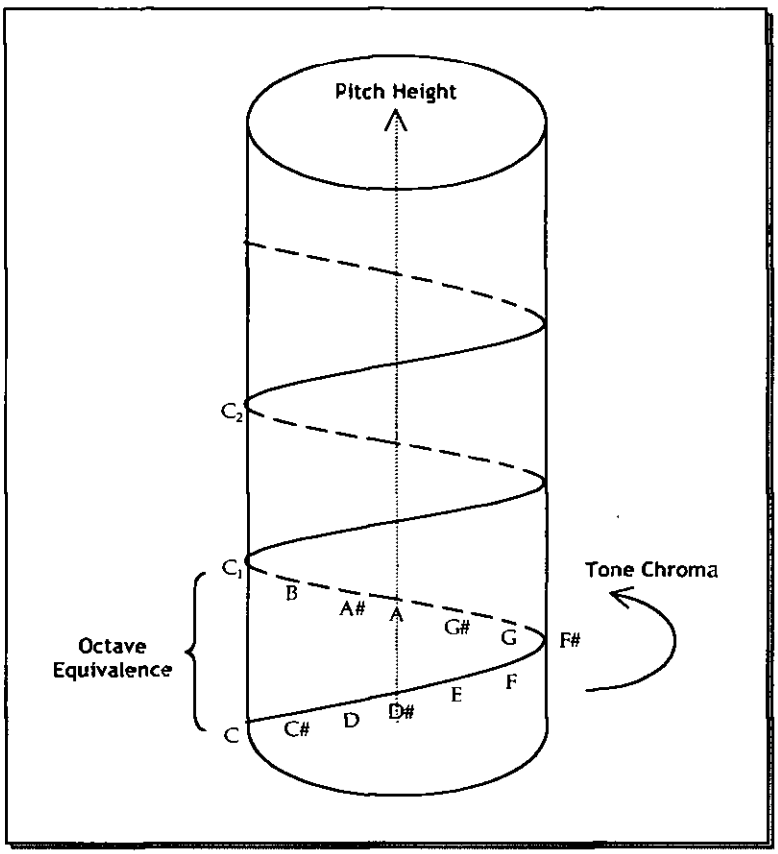


Figure 3.1: Two-dimensional representation of pitch based on Shepard (1999).

3.1.2 Loudness

Loudness is defined as "that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud" (Moore 1997). Loudness is our perception of sound intensity, i.e. the amount of air pressure variation in our ears. Sound intensity is a relative measure to an experimentally defined zero level (i.e. the lowest threshold of human hearing) and is expressed in decibels (dB). Attempts to define perceptually based measures of loudness have resulted in loudness scales such as the *phone* and *sone* (Stevens 1936). As in the case of the mel scale for pitch, the sone scale has been also criticised for being ambiguous and unreliable (Plomp and Rasch 1999).

For sinusoidal tones, the loudness level expressed in phones is equal to the sound intensity level in decibels only at the frequency of 1000 Hz. Fletcher and Munson (see Butler (1997)) investigated the relationship between phones and decibels by asking subjects to listen to a 1000 Hz standard tone together with a comparison tone with a lower or higher frequency and then adjust the loudness of the comparison tone until it matched the loudness of the standard tone. Although their results showed that phones and decibels are not equivalent throughout the frequency range, they are approximately equal in the range 300 Hz - 4000 Hz (see Figure 3.2). Furthermore, Butler (1997) argued that if complex tones were used as stimuli, the curves would flatten out on the left (low-frequency) side of the equal loudness curves. Based on this, it can be argued that in the case of complex tones which is of interest in sound synthesis (or for music purposes), sound intensity is a good indication of perceived loudness across the frequency range of musical interest. In addition, according to Plomp and Rasch (1999), "for the researcher, physical levels are the most precise reference and at the same time a rough indication of subjective loudness" (p. 99).

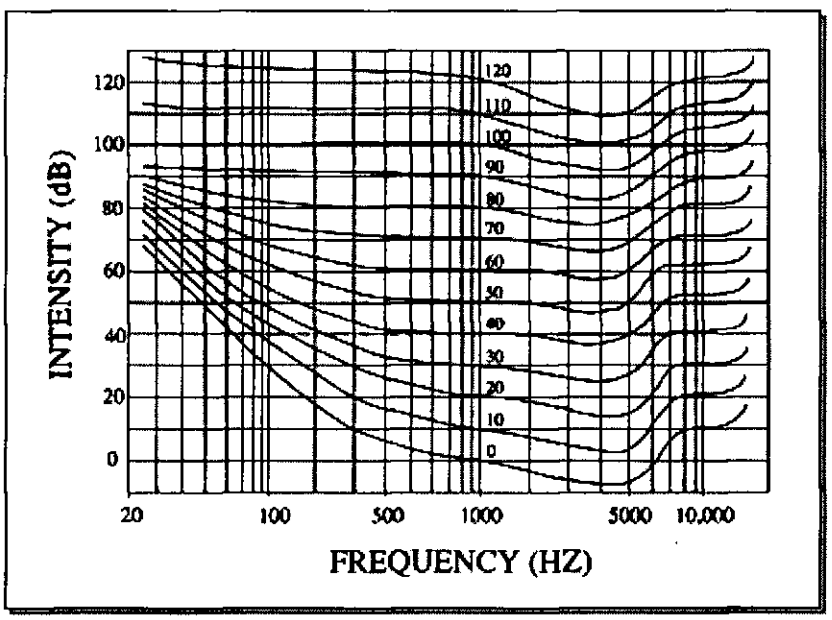


Figure 3.2: Equal-loudness curves after Fletcher and Munson (1933).

Finally, studies that measure the loudness of complex sounds have shown that loudness depends on the concept of *critical band*. The latter is "the bandwidth of frequencies within which energy is said to be summed to make up the loudness of a complex sound" (Slawson 1985). For example, if the frequency components of a complex sound with a given sound intensity fall within one critical band (the size of the bandwidth is usually 100 Hz -160 Hz), the sound is as loud as a pure tone of equal intensity at the centre frequency of the bandwidth (Moore 1997). However, if the frequency components are

spread over more than one critical band, loudness begins to increase and is assumed to be equal to the sum of loudness contributions of successive adjacent critical bands (Plomp 1976). As discussed in more detail in the next section, this concept of *loudness summation* has a fundamental role in the perception of timbre.

3.1.3 Timbre

One of the main advantages and great potentials of computer-based sound synthesis is that it provides access to the design of sounds that extend beyond traditional instruments, thus allowing composers to explore new sound spaces. It should be clarified that when we talk about the design of new sounds we refer primarily to the quality or *timbre* of the sounds.

Timbre has been defined by ASA (1960) as "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar". This definition has been strongly criticised for being too general and ill-defined (see Bregman (1990), Slawson (1985)). In fact, the term 'timbre' is used in a variety of contexts and it is extremely difficult to agree on a single definition. For example, timbre may refer to a class of musical instruments (e.g. string instruments as opposed to brass instruments), a particular instrument in this class (e.g. violin), a particular type of this instrument (e.g. Stradivarius), the quality of playing this instrument (e.g. bad or good quality), and so on. Therefore, any attempts to investigate the dimension of timbre should clarify which aspects of timbre are addressed.

In our thesis, timbre is defined as that perceptual attribute which pertains to the steady-state characteristics of sound. Although temporal characteristics are equally important and necessary for a complete description of timbre (Grey 1975), they are more related with the identification of sound sources and their intrinsic behaviour rather than the qualitative characteristics of timbre that are hidden in the steady-state spectrum of sounds.

There is general agreement in studies of auditory perception that timbre depends on certain characteristics of the sound spectra. This is primarily based on the assumption that the human auditory system consists of a number of bandpass filters that perform an analysis of the incoming sound (Moore 1997). A bandpass filter boosts frequencies that fall in a certain bandwidth while attenuating frequencies below and above the limits of the bandwidth. The spectrum of a bandpass filter is determined by the centre frequency of the resonance and its bandwidth. Under this scope, timbral characteristics are related to the relative levels produced by a sound in each of these filters or critical bands. For example, Plomp (1976) has showed the dependence of timbre upon the spectrum of sounds in a series of experiments that compared the differences in the perception of

sounds with the differences in their spectra. The sounds were vowel sounds and steady-state parts of sounds produced by nine musical instruments and their spectra were analysed in terms of the relative output levels in 15 (for instrument sounds) and 18 (for vowel sounds) 1/3-octave frequency bands. After plotting the differences, Plomp found a very close match, a result that led him to conclude that timbre is related to the relative level produced by a sound in each frequency band (or critical band). Although a critical band can be specified for every frequency in the audible frequency range, according to Moore (1997) the number of dimensions required to characterise timbre is limited by the number of critical bands required to cover the audible frequency range, i.e. a maximum of 37 dimensions.

It becomes evident from the above discussion that timbre, in contrast to pitch and loudness, is a multidimensional attribute of auditory perception. Many studies attempted to identify the prominent dimensions of timbre (for example Bismarck (1974a), Grey (1975), Ehresman and Wessel (1978), McAdams (1999)). These studies suggest that there is a limited number of dimensions on which every sound can be given a value, and that if two sounds have similar values on some dimension they are alike on that dimension even though they might be dissimilar on others. However, with very few exceptions, there is no agreement on the dimensions of timbre that these studies have proposed. This is mainly due to the different sets of sounds that were used as stimuli in the experiments (e.g. instrument tones as opposed to synthetic tones) and the different time portions of the sounds that were investigated (e.g. attack transients as opposed to steady states). As a result, the findings of these studies hold very well for the limited range of sounds that they refer to but they lack generality of application (Bregman 1990). McAdams *et al* (1995) speculated that there could be a set of common dimensions shared by all timbres while additional perceptual dimensions may be specific to certain timbres.

In the remainder of this section, we discuss dimensions of timbre that have been proposed in the related literature in an attempt to define a *perceptually-based* model of timbre. Dimensions of timbre that refer to temporal characteristics have been excluded from this discussion for the reasons stated earlier.

Sharpness

Sharpness (other terms include auditory brightness and spectral centroid) is the most prominent dimension of timbre suggested by the above-described studies. For pure tones, sharpness is determined by the fundamental frequency, i.e. the higher the fundamental frequency, the greater the sharpness. In the case of complex tones, the determining factors for sharpness are the upper limiting frequency and the way energy is distributed over the frequency spectrum, i.e. the higher the frequency location of the spectral envelope centroid, the greater the sharpness (Bismarck 1974a,b).

Compactness

Compactness (or tonalness) is a measure of a sound on a scale between complex tone and noise, i.e. the difference between discrete and continuous spectra (Bismarck 1974a). However, the formulation of such a scale has been proven difficult. Compactness is also related to the concept of periodicity, in the sense that periodic sounds are *tone-like* as opposed to aperiodic or *noise-like* sounds. Malloch (1997) suggested that *cepstrum* analysis (see Roads (1996)) as a method to measure the periodicity of a sound could also be used in the measurement of compactness.

Spectral Smoothness

Spectral smoothness is a dimension of timbre discussed in McAdams (1999). It describes the shape of the spectral envelope and it is a function of the degree of amplitude difference between adjacent partials in the spectrum of a complex tone. Therefore, large amplitude differences produce jagged envelopes, whereas smaller differences produce smoother envelopes.

Sensory Dissonance

Sensory dissonance (or roughness) is related to the phenomenon of *beats*. When two pure tones with very small difference in frequency are sounded together, then a distinct beating occurs that gives rise to a sensation of sensory dissonance (Sethares 1999). In a series of experiments with pairs of pure tones, Plomp (1976) found that sensory dissonance reaches its maximal point at approximately 1/4 of the relative critical bandwidth. For complex tones, sensory dissonance can be estimated as the sum of all the dissonances between all pairs of partials (see Sethares (1999), Hutchinson and Knopoff (1978)).

3.1.4 Evaluation of Auditory Dimensions

So far, we have discussed the properties of a number of prominent auditory dimensions, namely *pitch*, *loudness* and *timbre*. In this section, we present a critical evaluation of these dimensions based on the following criteria:

- **Empirical support.** This criterion tests whether a proposed dimension is supported by experimental work.

- **Independence.** This criterion tests whether a perceptual dimension is orthogonal, i.e. independent of changes in other dimensions or it is

somehow affected by these changes, in which case, the interactions between these dimensions should be taken into account.

- **Measurability**. This criterion tests the existence of concrete measurement methods for perceptual dimensions.

- **Synthesisability**. This is related to the criterion of measurability and refers to existing or potential models of synthesis algorithms that control perceptual dimensions.

The results of our critical evaluation are presented in the remainder of this section according to each of the above criteria.

Empirical Support

As far as the criterion of empirical support is concerned, the above-described dimensions of auditory perception are the results of rigorous empirical investigations involving musical and/or non-musical subjects within the limitations of their sound stimuli. Various experimental techniques have been employed, such as magnitude estimation, multidimensional scaling, and semantic differential. Therefore, we can be confident that sound synthesis tools that incorporate these dimensions are perceptually valid.

Independence

As far as the criterion of independence is concerned, although these dimensions are to a large extent independent of each other, various interactions between them have been reported in the related literature that can be outlined as follows:

- Interactions between the dimensions of tone chroma and pitch height. Although early empirical studies (e.g. Shepard (1964)) suggested that pitch height and tone chroma are orthogonal, the results of later studies described in Deutsch (1999) indicated that there is significant interaction between the two dimensions.

- Interactions between pitch and loudness. For example, the equal-loudness curves presented earlier in Figure 3.2 demonstrate that the loudness of a pure tone depends on its frequency. In addition, various studies (e.g. Olson (1967)) have shown that changing the loudness of a pure tone with constant frequency results in a perceptual change in pitch. However, these interactions are small and as Butler (1992) states "probably of negligible musical significance" (p. 209).

- The perception of timbre also depends on both pitch and loudness. For example, the perception of a particular timbre may be altered when sounded in different pitch registers. This interaction can be addressed by preserving the spectral envelope of the sound at different places along the frequency range (see Slawson (1985), Risset and Wessel (1999)). Timbre is also loudness-dependent, although this is primarily true for traditional music instruments. For example, playing an instrument with different dynamics can change dramatically the spectrum of the sound (Butler 1992). Therefore, this interaction should be taken into account when trying to imitate the sounds of traditional music instruments. From a broader sound synthesis perspective, it can be argued that this interaction is of limited significance.

- Interactions between dimensions of timbre. For example, changing the location of the frequency components in order to vary the amount of sensory dissonance for a particular sound may affect other dimensions such as the sharpness of the sound, since the location of the partials is taken into account when measuring these dimensions. In addition, dissonant and noise-like sounds do not evoke a clear perception of pitch. Therefore these interactions should be taken into account when designing tools for the independent control of these dimensions.

Measurability

Various measurement formulae have been proposed for the above-discussed auditory dimensions and the reader is referred to the individual studies discussed in the preceding sections for more detailed information on computational issues, although these are discussed in more detail in later parts of this thesis (§6.2.3).

Synthesisability

Finally, these dimensions have been primarily investigated in studies concerned with the analysis of sounds and there is no explicit discussion about the synthesisability of these dimensions. However, it seems feasible to create synthesis algorithms that are based on the measurement formulae, as discussed later in §6.2.3.

3.2 Conclusion

In this chapter, we reviewed the findings of various studies of auditory perception in order to identify prominent dimensions that can be incorporated in the design of sound design tools and in further investigations of the associations between auditory and visual dimensions.

The model of auditory perception that we propose comprises the dimensions of *pitch, loudness,* and *timbre.* At this stage of our research we decided to treat pitch as a one-dimensional attribute of sound as a result of the above-discussed interactions between pitch height and tone chroma. Furthermore, timbre is considered to comprise the dimensions of *sharpness, compactness,* and *sensory dissonance.* The dimension of *spectral smoothness* has been excluded from our present investigations because it appears to be related to visual characteristics of the sound's amplitude spectrum and it is not clear how it can be described in auditory terms. Our model is limited in the sense that it does not take into account other auditory dimensions such as duration, spatial characteristics, and temporal dimensions of timbre. These have been left for further work leading from the research described in this thesis (see §9.5).

In the next chapter, we begin our empirical investigations of how the dimensions incorporated in our model of auditory perception can be described in visual terms.

4

The Colour of Sound

This chapter is concerned with the application of colour in the visualisation of auditory dimensions for computer-based sound synthesis. As discussed in Chapter 2, existing approaches in this area have associated perceptual dimensions of colour with various auditory dimensions such as loudness, pitch, timbre, and stereo position. However, it was concluded that these associations were not empirically derived or validated. As a result there is a lack of a theoretical framework for auditory-visual associations that is based on empirical studies and that can be used in the formulation of design requirements for sound synthesis tools. In this chapter, we have investigated ways to overcome this limitation. We review existing studies in the field of colour perception. We describe and analyse the results of an experiment to investigate associations between a number of auditory and visual dimensions. Our results build upon those of previous experiments and provide a basis upon which more cognitively based sound design tools can be developed.

4.1 The Perception of Colour

Colour is one of the most fundamental aspects of visual perception. Although it seems natural to think about colour as a property of physical light, colour is a psychological phenomenon and refers to the sensation produced as a response to *visible* light transduced by our visual system. Visible light is defined as the electromagnetic spectra with wavelengths between 400 and 700 nanometers (Fortner and Meyer 1997).

According to current theories of colour vision, our perception of colour is based on a three-stage process. In the first stage, colour is the result of mixing the output of three different kinds of sensors in the eye (*red* (R), *green* (G), and *blue* (B)). We can perceive thousands of different colours that are produced by the differing ratios among the three sensors. These RGB outputs are then translated in the second stage to produce signals along three *opponent* channels: *red-green*, *yellow-blue*, and *black-white*. This is the basic premise of Hering's opponent colour theory (see Fortner and Meyer 1997) according to which our colour judgements are related to the colours' positions on the three opponent scales. Finally, the third stage translates the opponent signals into the following three perceptual dimensions of colour (see also Figure 4.1):

- **Hue**. This is the dominant wavelength in the power spectrum of a colour and in everyday terms it refers to the quality of the colour (for example redness, blueness, and so on).

- **Saturation**. This is the degree to which hue is perceived to be present in a colour. A *strong* saturated colour means that a very small amount of light of

different wavelengths is mixed in with a particular hue, whereas *weak* desaturated colours contain more wavelengths.

- **Lightness** or **Brightness**. This is our perception of how light or dark a colour appears. Both lightness and brightness are perceptual dimensions related to the physical attribute of luminance, however lightness is used to describe light that is reflected from an object whereas brightness is used for self-luminous objects such as a light bulb, or a computer screen.

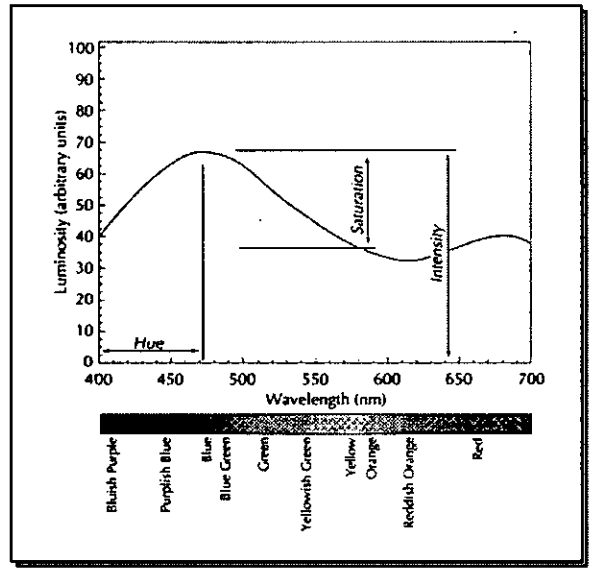


Figure 4.1: Conceptual view of calculating hue, saturation, and intensity from an arbitrary spectrum. In this figure of an example light spectrum, hue is around 472 nm (blue), intensity is around 67 (arbitrary units), and saturation is around 36/67 = 54% (after Fortner and Meyer (1997)).

An important characteristic of the above-discussed model of colour vision is that colour can be *completely* described at any stage by only three dimensions. As discussed briefly in §2.1.2 a formal method of representing the visual dimensions of colour is that of arranging colours in a *colour space* (Jackson *et al* 1994). For example, colours in the *RGB* colour space are described in terms of their values along the red, green, and blue channels. The RGB space is based on an additive mechanism by which the individual contributions of each channel are added to produce a particular colour sensation (Foley *et al* 1994). However, the RGB space has been described as *hardware-oriented* because the RGB values are not intuitive descriptors of colour (Foley *et al* 1994).

In order to address the limitations of hardware-oriented colour spaces, various *user-oriented* (or perceptually based) colour spaces have been proposed such as the HSV space (Smith 1978). The latter is based on the perceptual dimensions of hue (H), saturation (S),

and value (V), where value is the equivalent of brightness and this is why the model is also known as HSB, with *B* for brightness (Foley *et al* 1994). For consistency with our previous discussion, the term HSB will be used in the remainder of this thesis. The HSB colour space can be illustrated as the hexcone shown in Figure 4.2. Brightness is represented on the vertical axis. Hue is measured by the angle around the vertical axis, and by convention red is at 0°, yellow at 60°, and so on. Saturation is the distance of a point in this space from the brightness axis.

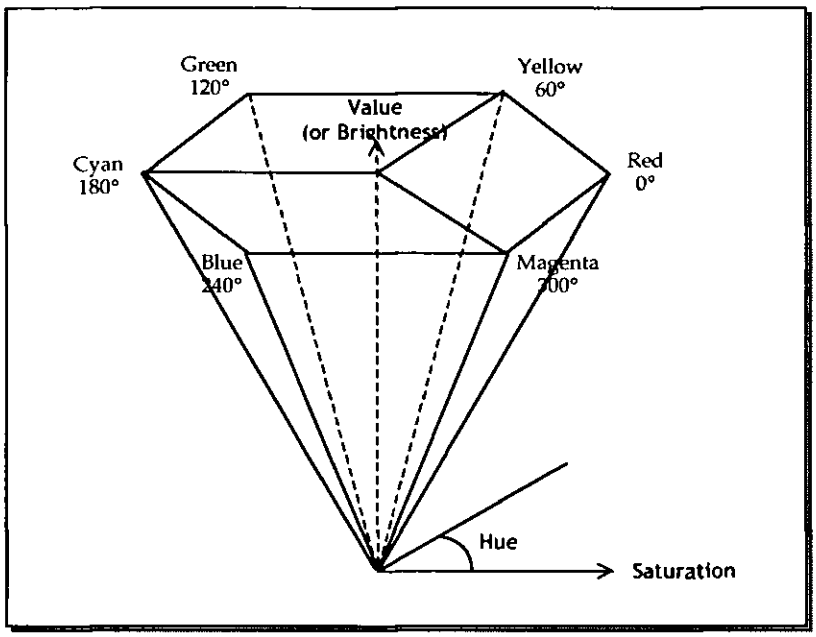


Figure 4.2: The HSB colour space (see also Foley *et al* (1994)).

The HSB colour space has been widely used for the interactive specification of colour in computer systems, where users can directly manipulate the three perceptual dimensions of hue, saturation, and brightness in order to specify a desired colour. However, a limitation of the HSB colour space is that it is not perceptually uniform, i.e. equal steps in any of the three dimensions do not correspond to equally perceived changes (Fortner and Meyer 1997).

4.2 Colour, Pitch and Loudness

In Chapter 3 we derived a model of auditory perception that comprises the perceptual dimensions of pitch, loudness, and timbre. The question that arises is which of these dimensions can be associated with perceptual dimensions of colour. For example, it is tempting to draw an analogy between timbre and colour since they are both multidimensional phenomena and can be described by a small number of dimensions. In

fact, such analogies have inspired timbre research studies that attempt to describe timbre by three dimensions (e.g. Pollard (1982)). However, although colour can be completely described by three dimensions, this is clearly not the case for timbre as discussed in §3.1.3. Therefore, it seems that the three perceptual dimensions of colour are not enough for a complete description of timbre. Colour dimensions could be more appropriate for the visualisation of pitch and/or loudness since associations between these perceptual dimensions have been to some extent empirically supported in studies of synaesthesia and cross-modal associations (see §2.1.3). To this end, an experiment was designed to investigate the matching of pure tones with colours. The main objective was to provide results to help answer the following questions:

- To what extent can a colour model based on hue, saturation, and brightness provide a useful metaphor to describe loudness and pitch?

- Which of these colour dimensions are associated with loudness and pitch?

4.2.1 Method

Experimental Design

We chose a between-subjects design, where each participant performed a series of sound-colour association tasks for a series of eleven sound sequences drawn from three different sequence orders as described later.

Subjects

We had twenty-four subjects in total and all were given a screening questionnaire about their experience in both traditional and computer music (see §A.1 for a copy of the questionnaire). The exact composition of the twenty-four subjects was:

- Twelve undergraduate students studying sonic arts (11) or other fields (1) with average music experience.

- Five individuals with a great deal of computer music experience.

- Seven individuals with no musical background.

Although the music experience of subjects was recorded, the experiment was not designed to test any effects of musical experience on subjects' responses. This decision was primarily based on the nature of the auditory dimensions involved in this

experiment. It can be argued that both pitch and loudness are familiar dimensions and their perception is to a large extent independent of musical experience.

Subjects were randomly assigned into three groups of eight (one group for each series of eleven sound sequences) and screened with an Ishihara colour plates test to detect colour vision deficiencies (one subject failed the Ishihara test). The purpose of the colour vision test was not to disqualify subjects but to test any effect(s) of colour blindness on subjects' colour selections.

Apparatus and Stimuli

A prototype computer application was designed in MacProlog32 (LPA 1998) for use in this experiment comprising a custom colour palette and three series of eleven sound sequences.

The design of the colour palette was based on a computer implementation of the previously described HSB colour model. We selected six hues: *red* (R), *yellow* (Y), *green* (G), *cyan* (C), *blue* (B), *magenta* (M) — in other words the three primary (RGB) and three secondary (YCM) hues. The saturation and brightness levels were subdivided into six equal steps thus producing thirty-six different saturation-brightness combinations for each hue (6×36 = 216 colours in total). The HSV values were then translated into their RGB equivalents encoded in the MacProlog32 programming environment in order to display the custom colour palette (see Figure 4.3 and Colour Plate E.1).

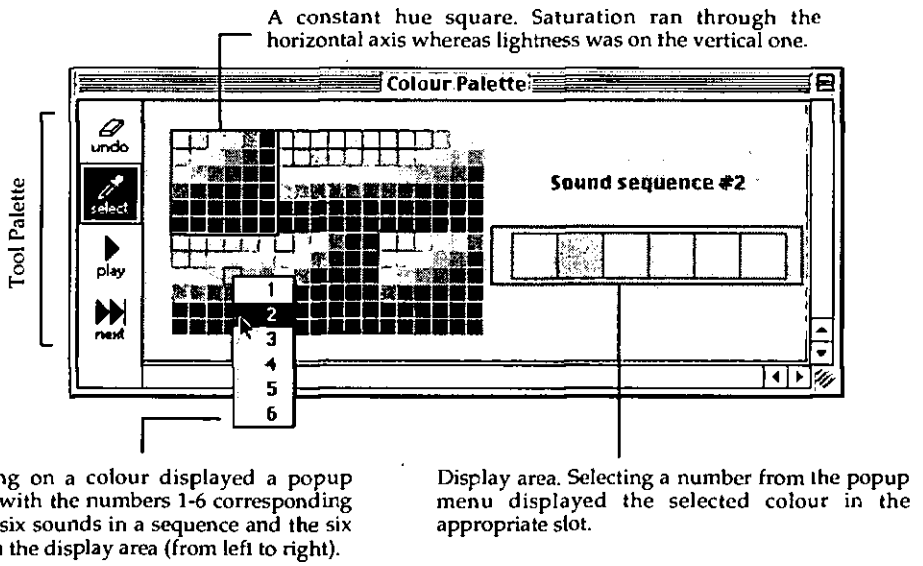


Figure 4.3: The prototype application used in the colour-sound experiment. See also Colour Plate E.1.

The auditory dimensions under examination in this experiment were loudness and pitch. We used pure tones, i.e. sounds with a single sinusoidal frequency component, in order to neutralise the effect of timbral richness (sound complexity) on subjects' responses. All sequences consisted of six sounds whose frequency content was a single fundamental frequency. The individual tones were designed with *PowerSynthesiser* (Russell 1995), a computer application for the design of psychoacoustical experiments involving sound. Although *PowerSynthesiser* provides adequate control over frequency and amplitude, mention should be made that these objective physical properties of sound are closely related to pitch and loudness, which in contrast are subjective perceptual measures as discussed in §3.1 and §3.2 respectively.

Three series of eleven sound sequences were designed — one series for each of the following frequency ranges: 110 Hz - 220 Hz (Low), 440 Hz - 880 Hz (Mid), and 1760 Hz - 3520 Hz (High). It can be seen that in each case, these frequency ranges represent *one-octave* intervals, i.e. the lowest and highest components in each of these frequency ranges are in a ratio of 2:1. Each frequency range was subdivided in six equal steps, which formed the content of the sound sequences used in this experiment. The sequences were designed and classified according to their level of complexity. The content of each sound sequence is depicted in Table 4.1.

Complexity ->	1				2		3				4
Sequence ->	1	2	3	4	5	6	7	8	9	10	11
Loudness	↑	↓	—	—	×	—	↑	↓	↑	↓	×
Pitch	—	—	↑	↓	—	×	↑	↓	↓	↑	×

—: Constant, ↑: Ascending, ↓: Descending, ×: Non-monotonic

Table 4.1: The sound sequences used in the colour-sound experiment.

The first complexity level comprised sequences where tones were either increasing or decreasing monotonically in one auditory dimension while keeping the other constant. The second level was an extension to the previous case with tones varying in a non-monotonic way. The third level incorporated sequences with both loudness and pitch varying simultaneously either in the same or opposite monotonic direction. Finally, the fourth level extended the previous case with non-monotonic variation of loudness and pitch. It should be mentioned that during the experiment, sequences were not introduced in the same order as depicted in Table 4.1. Instead, they were shuffled and their order was the same for all subjects within a subject group but in all cases sequences started with low complexity and progressed to higher complexity.

Experimental Task

The experimenter demonstrated how to use the prototype application shown in Figure 4.3. This was followed by a short practice period of one tone sequence. The practice sequence was part of the series but reintroduced later in the experiment. The experimental task was: for the current sequence of six tones to create a sequence of six corresponding colours. Subjects could listen to the current sequence as many times as they wished, at any point during the task. Each subject completed the task for eleven sequences. Subjects performed the experiment at their own pace and times ranged from thirty to forty-five minutes. The experimenter was present throughout the experiment recording observations that formed the basis for post experiment interviews with subjects. Finally, a data collection program logged colour selections in terms of hue, saturation, and brightness, as well as completion time per sequence.

Experimental Environment

The experiment was conducted in a room with normal 'office' lighting and sounds were presented binaurally through headphones. Due to hardware limitations the experiment was designed and run on an Apple PowerMac personal computer capable of representing thousands of colours, a limitation that in some cases produced less uniform colour variation in our Colour Palette. Subjects sat approximately 80cm away from the computer screen and the components of the interface were sized for comfortable viewing and manipulation at that distance.

4.2.2 Analysis of Results

We now present the results obtained from the above-described experiment. The presentation is based on a qualitative method supported by quantitative data. The major qualitative variable is the colour selection strategy followed by subjects. With three colour dimensions there are $2^3 = 8$ possible strategies. Tables 4.2 - 4.9 show the results after the processing of the raw data obtained from subjects' colour selections for each colour strategy. In the case of colour strategies that involved variation in a single colour dimension, the results show subjects' selections that matched at least four out of the six possible steps in the corresponding dimension. Furthermore, Figures 4.4 - 4.11 are based on the sequences created by all subjects and display average levels for each colour dimension relative to pitch and/or loudness. These figures should be interpreted as showing (a) the trend in subjects' responses, i.e. how close the average subjects' responses were to the desired variation levels, and (b) an indication of the correlation (positive, negative, none) between subjects' sequences and the sequence stimuli.

Table 4.2 shows the overall results for sequences where the only varying (monotonic and non-monotonic) auditory attribute was loudness. In single dimension terms (i.e. varying one dimension while keeping the remaining dimensions constant), subjects varied hue in 1/72, saturation in 38/72, and brightness in 9/72 sequences. Furthermore, hue remained constant in 55/72, saturation in 15/72, and brightness in 49/72 sequences. These results suggest that the majority of subjects from all three groups varied the saturation level while keeping hue and brightness at constant levels. This correlation between saturation and loudness can also be seen in Figure 4.4. Quiet sounds were associated with weak colours while louder sounds evoked stronger colour selections.

Table 4.3 shows the overall results for sequences where the only varying (monotonic and non-monotonic) auditory attribute was pitch. In this case the results are not as clear as for loudness. There was no dominant colour selection strategy as well as a high number of selections that involved variation in all three perceptual dimensions of colour. The results for strategies that involved variation in only one colour dimension show that subjects varied hue in 10/72 sequences, saturation in 15/75, and brightness in 11/72. However, these figures are relatively small to support safe conclusions, although Figure 4.5 hints at a possible pitch-brightness association. Furthermore, we can suggest that hue does not seem to have any immediate role in colour selections since hue remained constant in 41/72 sequences, saturation in 23/72, and brightness in 34/72.

Loudness			Monotonic and Non-monotonic			
Selection strategy			Frequency range			Total
H	S	V	Low	Mid	High	
-	-	-	0	3	0	3
✓	-	-	1	0	0	1
-	✓	-	13	14	11	38
-	-	✓	5	1	3	9
✓	✓	-	1	3	3	7
✓	-	✓	0	0	2	2
-	✓	✓	2	1	2	5
✓	✓	✓	2	2	3	7
Total sequences			24	24	24	72

Table 4.2: Overall results for sequences with varying loudness (monotonic and non-monotonic) and constant pitch.

Pitch			Monotonic and Non-monotonic			
Selection strategy			Frequency range			Total
H	S	V	Low	Mid	High	
-	-	-	1	0	0	1
✓	-	-	5	3	2	10
-	✓	-	4	4	7	15
-	-	✓	6	3	2	11
✓	✓	-	3	2	3	8
✓	-	✓	1	0	0	1
-	✓	✓	0	11	3	14
✓	✓	✓	4	1	7	12
Total sequences			24	24	24	72

Table 4.3: Overall results for sequences with varying pitch (monotonic and non-monotonic) and constant loudness.

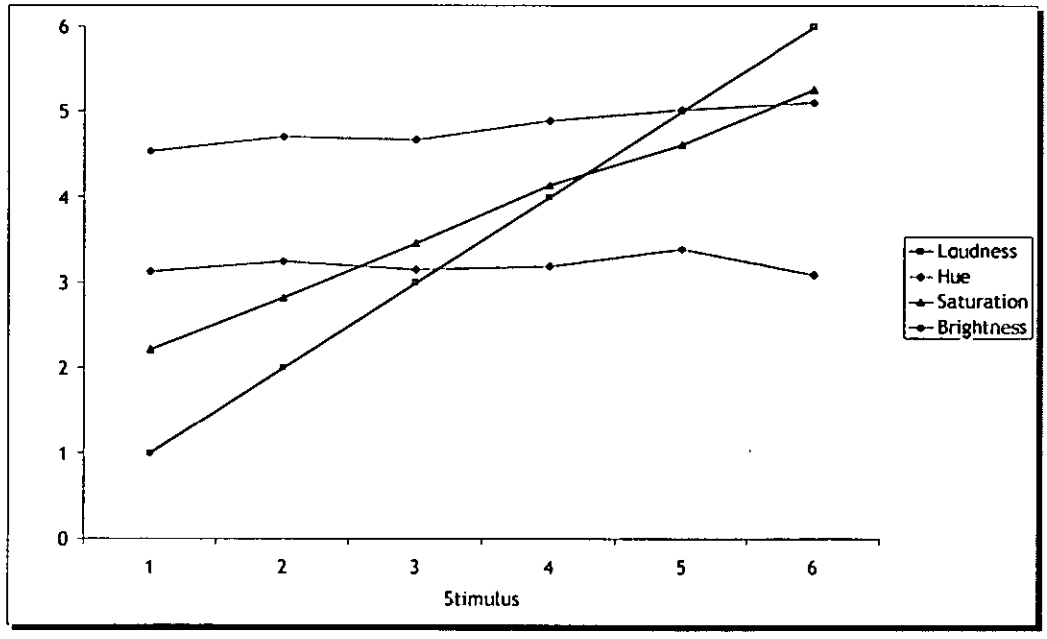


Figure 4.4: Average subjects' responses in each colour dimension for sound sequences with monotonic and non-monotonic variation in loudness.

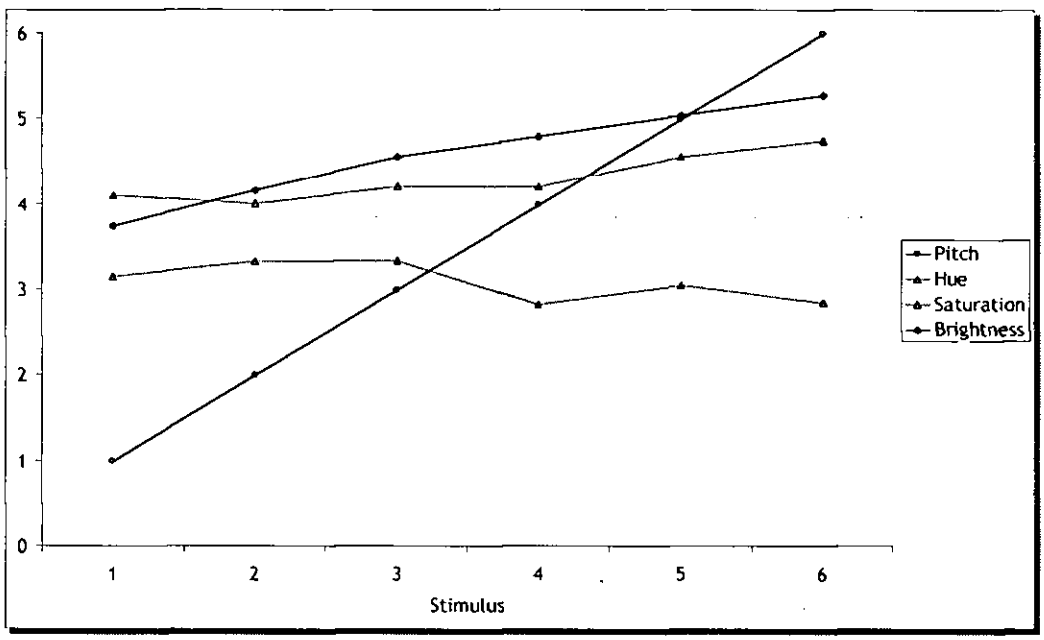


Figure 4.5: Average subjects' responses in each colour dimension for sound sequences with monotonic and non-monotonic variation in pitch.

Table 4.4 shows the overall results for sequences where both loudness and pitch were varying simultaneously in a positively correlated fashion. Here, the dominant colour selection strategy (17/48 sequences) was to vary both saturation and brightness while hue remained constant (33/48 sequences). This supports the point made before that saturation and brightness were the key dimensions for differences in loudness and pitch. This can be also seen in Figure 4.6. However, since both loudness and pitch follow the same pattern (either soft-loud/low-high or loud-soft/high-low), we cannot immediately tell which of the two corresponds to which auditory dimension. In order to address this issue we examined the subjects' responses for the third level of sequence complexity, i.e. sequences with negatively correlated levels of loudness and pitch. Tables 4.5 and 4.6 contain the results for these two cases (see also Figures 4.7 and 4.8). The results suggest that in both cases saturation and brightness were again the varying colour dimensions (hue remained constant in 17/24 and 16/24 sequences respectively). Table 4.7 breaks down the results for this colour selection strategy in sequences where loudness descended monotonically from loud to soft and pitch followed the reverse pattern. The dominant strategy was to decrease saturation and increase brightness levels. Based on these results we can suggest that saturation and brightness were associated with loudness and pitch respectively. However, this would hold only if the same applied to sequences where pitch descended monotonically from high to low and loudness followed the reverse pattern. This is clearly demonstrated from the results shown in Table 4.8.

Loudness - Pitch			Same monotonic variation			
Selection strategy			Frequency range			Total
H	S	V	Low	Mid	High	
-	-	-	0	0	1	1
✓	-	-	1	0	5	6
-	✓	-	1	5	4	10
-	-	✓	3	0	2	5
✓	✓	-	1	0	1	2
✓	-	✓	1	0	0	1
-	✓	✓	7	8	2	17
✓	✓	✓	2	3	1	6
Total sequences			16	16	16	48

Table 4.4: Overall results for sequences with the same monotonic variation in both loudness and pitch.

Loudness - Pitch			Descending Loudness - Ascending Pitch			
Selection strategy			Frequency range			Total
H	S	V	Low	Mid	High	
-	-	-	0	0	1	1
✓	-	-	0	0	2	2
-	✓	-	1	2	1	5
-	-	✓	0	0	1	1
✓	✓	-	0	1	1	2
✓	-	✓	1	0	0	1
-	✓	✓	5	5	1	11
✓	✓	✓	1	0	1	2
Total sequences			8	8	8	24

Table 4.5: Overall results for sequences with descending loudness and ascending pitch.

Loudness - Pitch			Ascending Loudness - Descending Pitch			
Selection strategy			Frequency range			Total
H	S	V	Low	Mid	High	
-	-	-	0	0	0	0
✓	-	-	0	1	2	3
-	✓	-	1	1	1	3
-	-	✓	1	0	2	3
✓	✓	-	0	1	1	2
✓	-	✓	1	0	0	1
-	✓	✓	4	4	2	10
✓	✓	✓	1	1	0	2
Total sequences			8	8	8	24

Table 4.6: Overall results for sequences with ascending loudness and descending pitch.

Loudness - Pitch		Descending Loudness - Ascending Pitch			
Selection strategy		Frequency range			Total
S	V	Low	Mid	High	
↑	↑	1	1	0	2
↑	↓	1	1	0	2
↓	↓	0	0	0	0
↓	↑	3	3	1	7
Total sequences		5	5	1	11

Table 4.7: Breakdown of results for the saturation-brightness strategy based on the results in Table 4.5.

Loudness - Pitch		Ascending Loudness - Descending Pitch			
Selection strategy		Frequency range			Total
S	V	Low	Mid	High	
↑	↑	0	0	0	0
↑	↓	4	3	0	7
↓	↓	0	1	2	3
↓	↑	0	0	0	0
Total sequences		4	4	2	10

Table 4.8: Breakdown of results for the saturation-brightness strategy based on the results in Table 4.6.

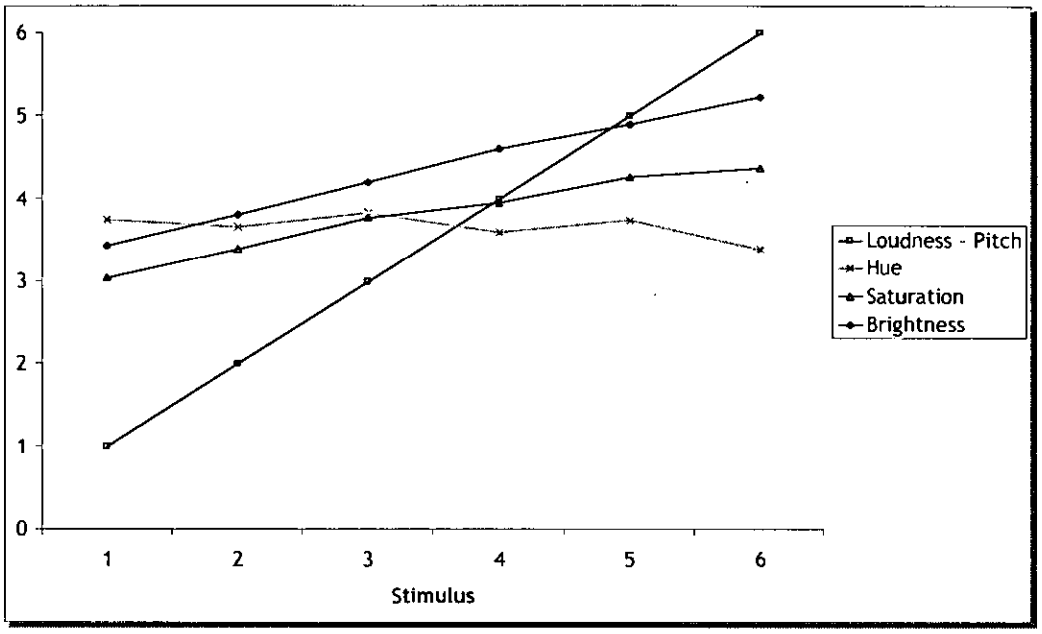


Figure 4.6: Average subjects' responses in each colour dimension for sound sequences with the same monotonic variation in loudness and pitch.

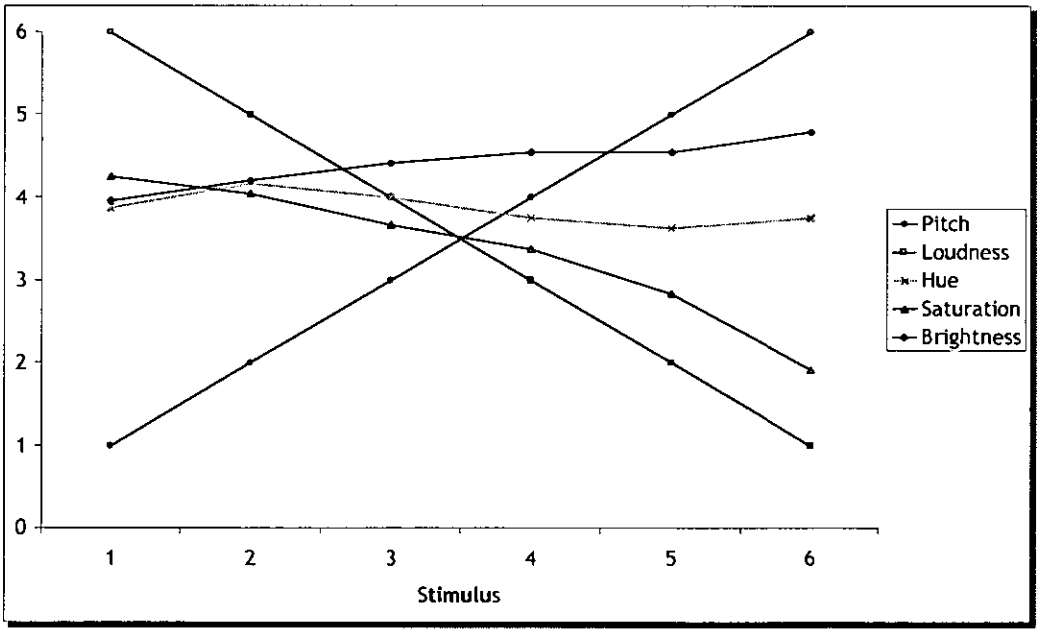


Figure 4.7: Average subjects' responses in each colour dimension for sound sequences descending in loudness and ascending in pitch.

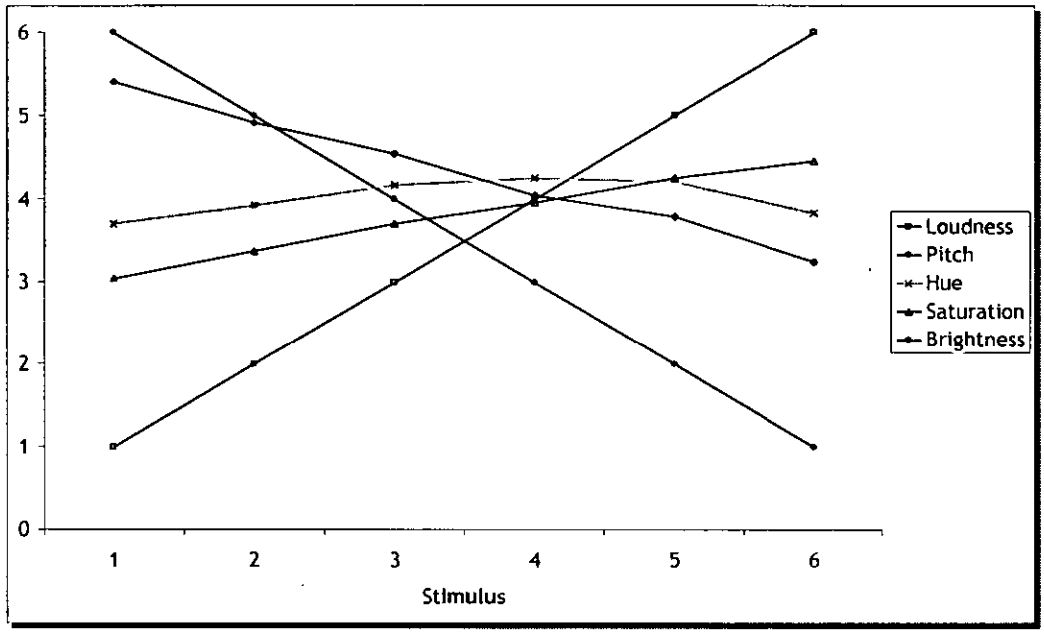
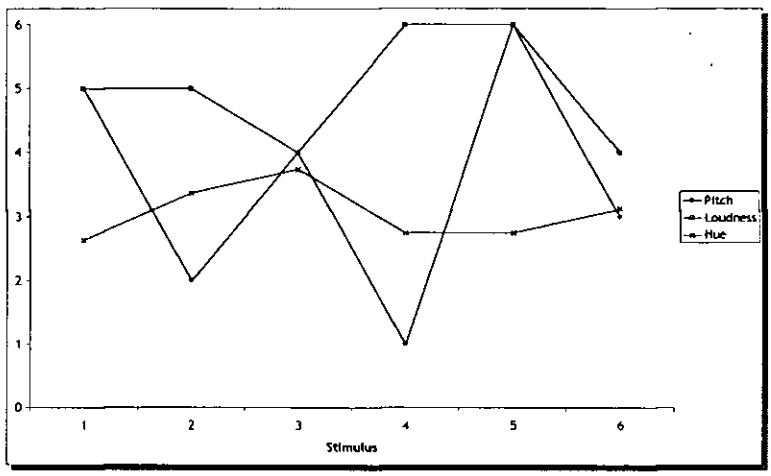


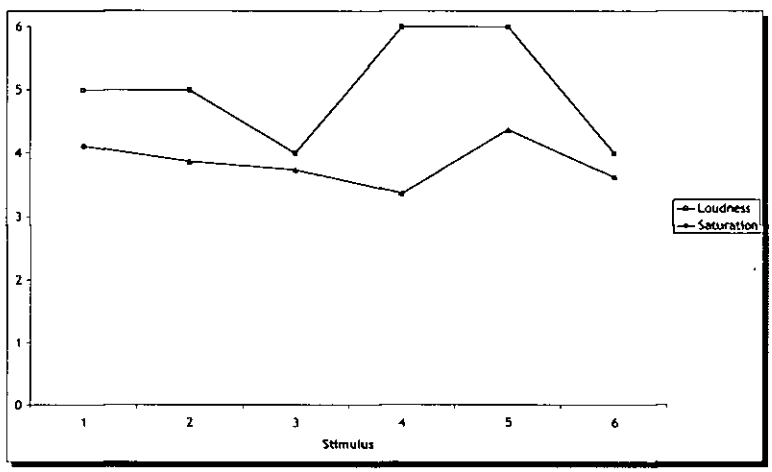
Figure 4.8: Average subjects' responses in each colour dimension for sound sequences ascending in loudness and descending in pitch.

The results for the fourth level of complexity are shown in Table 4.9. Once again, subjects varied saturation and brightness as a response to the non-monotonic simultaneous variation in both pitch and loudness. As can be seen in Figures 4.9, 4.10, and 4.11, the variations in saturation and brightness seem to match the variations in loudness and pitch respectively, however, the complexity of the sequences clearly affected the accuracy of colour selections.

Loudness - Pitch			Non-monotonic			
Selection strategy			Frequency range			Total
H	S	V	Low	Mid	High	
-	-	-	1	0	0	1
✓	-	-	1	0	0	1
-	✓	-	0	1	2	3
-	-	✓	1	0	0	1
✓	✓	-	0	0	3	3
✓	-	✓	0	0	0	0
-	✓	✓	2	4	3	9
✓	✓	✓	3	3	0	6
Total sequences			8	8	8	24

Table 4.9: Overall results for sequences with non-monotonic variation in both loudness and pitch.





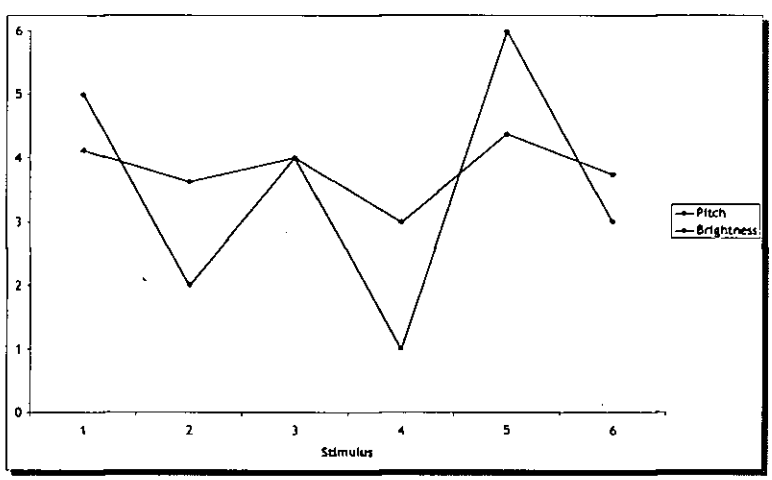


Figure 4.9: Average subjects' responses in each colour dimension for sound sequences with non-monotonic variation of loudness and pitch (low-frequency range). The top figure shows the results for hue and pitch/loudness. The middle and bottom figures show the results for saturation-loudness and pitch-brightness.

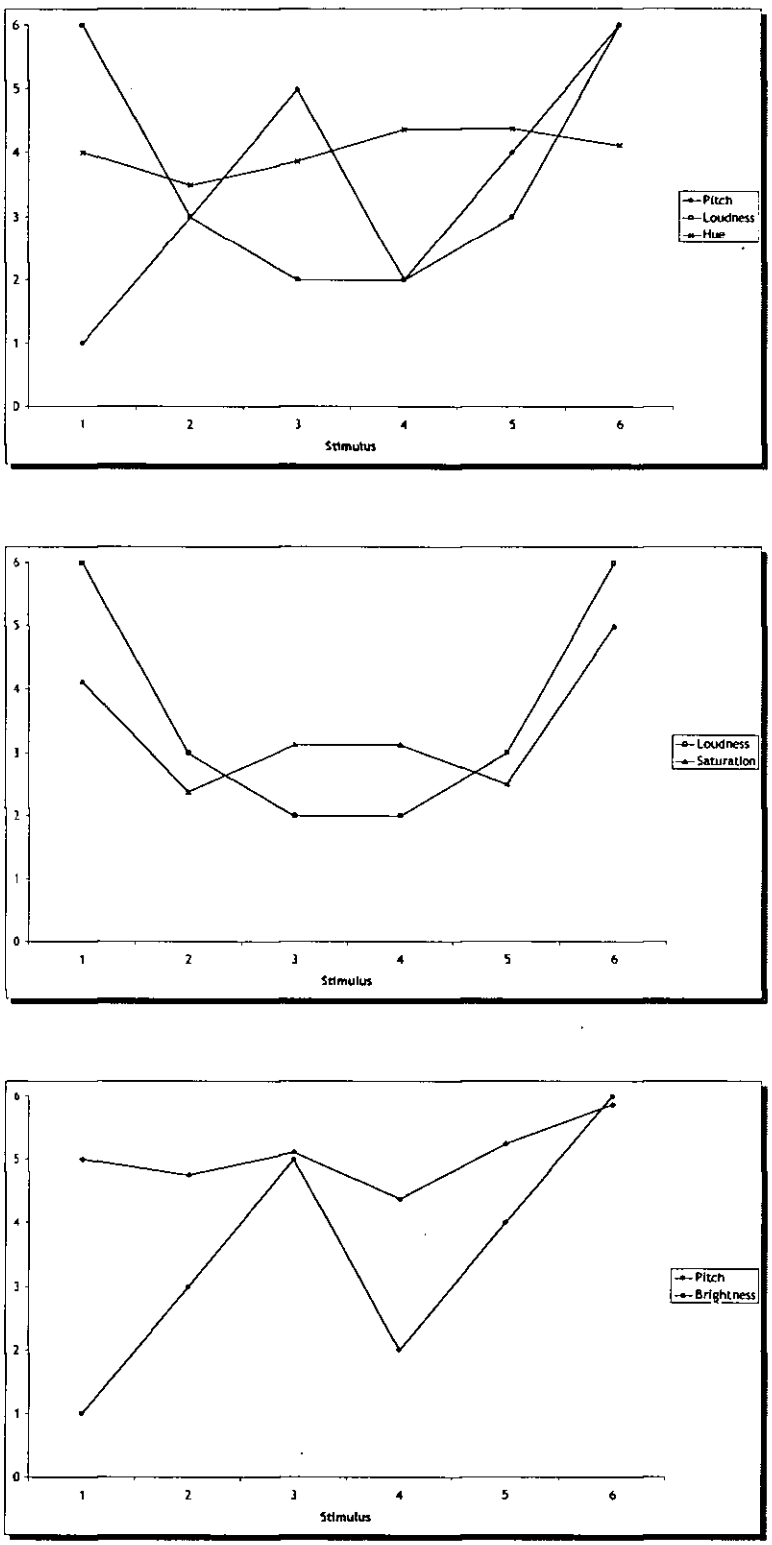
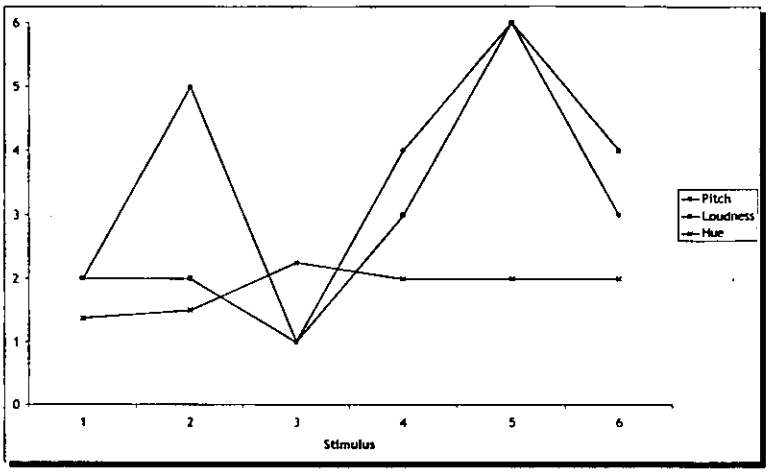
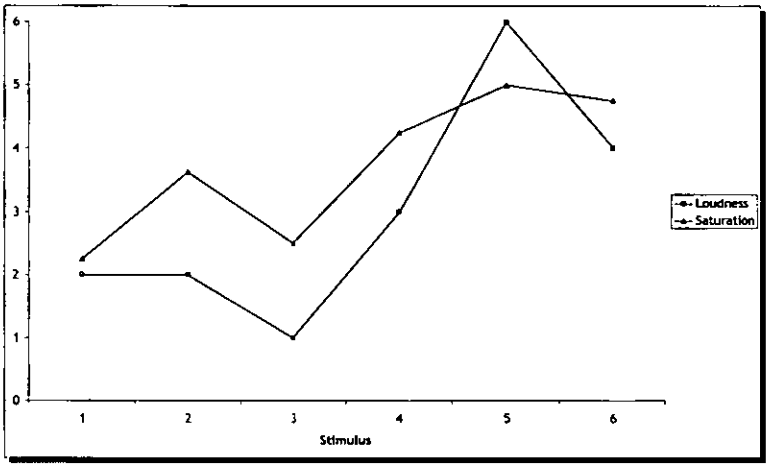


Figure 4.10: Average subjects' responses in each colour dimension for sound sequences with non-monotonic variation of loudness and pitch (mid-frequency range). The top figure shows the results for hue and pitch/loudness. The middle and bottom figures show the results for saturation-loudness and pitch-brightness.





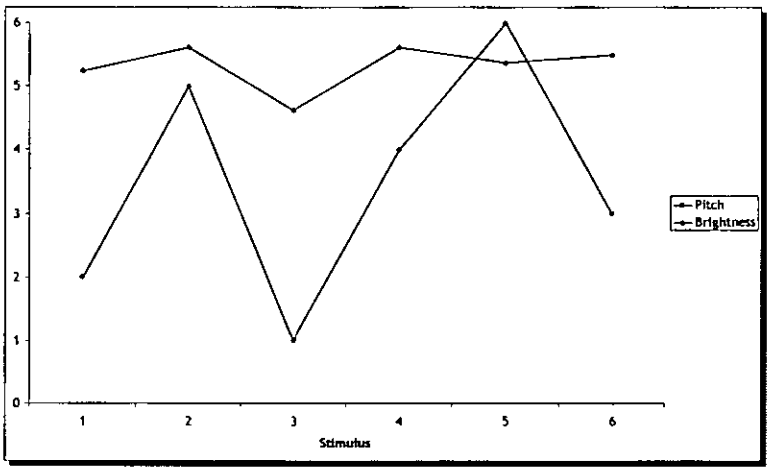


Figure 4.11: Average subjects' responses in each colour dimension for sound sequences with non-monotonic variation of loudness and pitch (high-frequency range). The top figure shows the results for hue and pitch/loudness. The middle and bottom figures show the results for saturation-loudness and pitch-brightness.

Finally, we examined the colour selection strategy that involved no variation in any colour dimension, i.e. subjects selected the same colour for all the tones in the sequence despite the variation in loudness and/or pitch. These results are shown in the first row of figures for the tables discussed above (except Tables 4.7 and 4.8). Summing up these figures results in six such cases which, surprisingly, belong only to the subject that failed the Ishihara test. However, since there was only one colour-blind subject, the above observation has no significant statistical value. Furthermore, this subject reported during the post-experiment interview that no variation was perceived in the majority of the sound sequences. Based on this latter response we can speculate that the particular subject could have been tone-deaf.

As previously mentioned the vast majority of subjects kept hue at constant levels. It is of further interest to examine to what extent these levels were related to sound characteristics. In Table 4.10 we have compiled the results for all the sequences where hue remained constant. The hues are organised in pairs and in the same order as they appear in the HSB colour space. For sequences of low frequency tones, chosen hues appear to fall most often in the *blue-magenta* region. For sequences of middle frequency tones, chosen hues appear to fall most often in the *green-cyan* region. Finally, for high frequency tones, chosen hues appear to fall most often in the *red-yellow* region. Therefore, although hue does not seem to have any immediate effect on colour selections, there seems to be an effect in terms of general frequency ranges. This means that subjects might have associated hue with certain frequency ranges and varied brightness with the various frequencies that fall in those ranges. However, these results are not as clear as for the dimensions of brightness and saturation.

Hue Pairs	Frequency range			Total
	Low	Mid	High	
Red-Yellow	16	20	28	64
Green-Cyan	20	28	10	57
Blue-Magenta	26	19	8	51
Total seque nces	62	67	46	175

Table 4.10: Compiled results for all the sequences where hue remained constant.

Finally, there is a noticeable difference in the results presented in Tables 4.2 - 4.9 for sound sequences drawn from the high frequency range. In the majority of these cases the obtained results were less clear than for other frequency ranges. Although, this can be attributed to the particular frequency range, it is our speculation that this is a result of the subjects that were presented with these sequences. In more detail, high-frequency

sequences were mainly tested with our third subject group which was primarily composed by non-music subjects, despite the fact that subjects were randomly assigned into groups. Therefore, the obtained results suggest that musical experience may have an important role even in the case of familiar auditory dimensions such as pitch and loudness.

4.3 Conclusion

Based on the above analysis and discussion we can argue that the loudness and pitch of pure tones can be predicted by saturation and brightness respectively as shown in Figure 4.12. In general, quiet tones were associated with low levels of saturation and louder tones with increasing levels of saturation. Furthermore, low-pitched tones evoked dark (low levels of brightness) colour selections while high-pitched tones were associated with lighter colours. Hue was not found to have any immediate association with pitch or loudness. However the experimental results suggest an association between hue and certain sound frequency ranges. Finally, our experimental design suggests that the use of a three-dimensional colour space can provide a more useful framework for the investigation of auditory-visual associations than the single dimension scales used by previous studies (for example Marks (1975)).

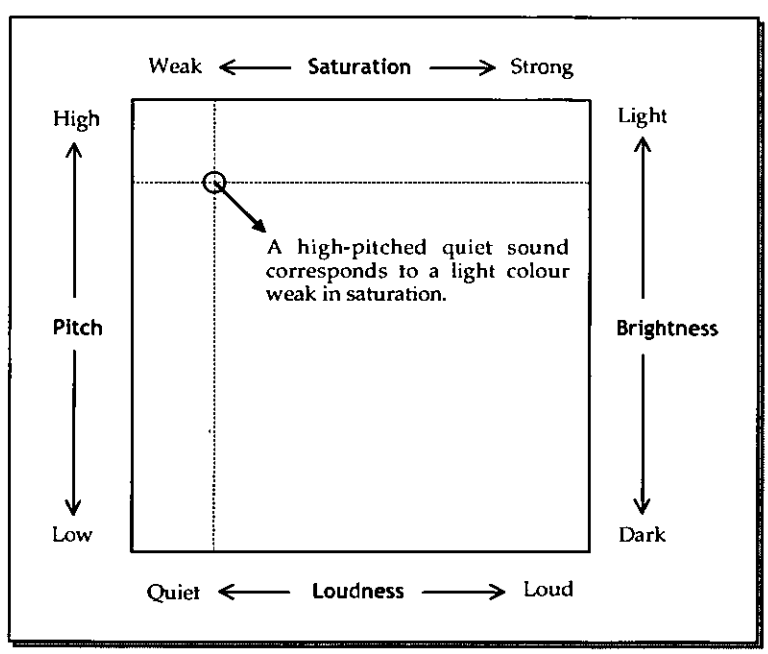


Figure 4.12: Proposed space for the associations between pitch-brightness and loudness-saturation.

However, there are various experimental design flaws that should be taken into account when interpreting our results. First, since the order of the sequences was the same for all the subjects in the same group, ordering effects were not controlled for. Second, it appears that our initial assumption that musical experience would have no effect on subjects' responses was wrong since the results obtained from non-music subjects were not as clear as those from music subjects. In addition, the arrangement of the colour chips in our custom colour palette might have made it easier for subjects to change the brightness and saturation levels while focusing on one hue square of the palette, thus disfavouring the selection of different hues. Finally, sound sequences were drawn from all frequency ranges, so an alternative design would have been to assign subject groups to only one frequency range in order to test the effect of frequency ranges. This could have also given more insight on the role of hue in colour-sound associations. Nevertheless, it is our belief that the above limitations did not influence positively subjects' responses; on the contrary, we believe that these limitations were the primary sources of ambiguity in certain cases.

5

The Texture of Sound

In the previous chapter, we argued that a colour space comprising the dimensions of brightness and saturation could be used for the visualisation of auditory pitch and loudness respectively. Our study involved the use of pure tones as experimental stimuli in order to neutralise the effect of timbral richness. A natural extension of our research is the empirical investigation of the associations between dimensions of timbre and dimensions of other visual percepts such as texture, shape, motion, and spatial dimensions among others.

As discussed in §3.1.3, the perception of timbre is a highly complex and multidimensional phenomenon. Recently, visual texture has been used effectively in studies such as the visualisation of multidimensional data sets (see for example Ware and Knight (1992), Healey and Enns (1998)), and visual information retrieval (e.g. IBM's Query By Image Content (QBIC) project (Niblack *et al* 1993), Gupta and Jain (1997), Bimbo (1999)). The motivation behind these studies is to exploit the sensitivity of the human visual system to texture in order to overcome the limitations inherent in the display of multidimensional data and provide more intuitive ways for searching and retrieving information from image and video databases (Rao 1996).

In this chapter, we begin with a review of existing studies in the perception of visual texture in order to define an appropriate set of perceptual dimensions of visual texture that can be incorporated in further investigations (§5.1). In addition, we present the results of an experiment that we designed and conducted for the empirical investigation of the cognitive associations between timbre and visual texture (§5.2).

5.1 The Perception of Visual Texture

Texture is a property that can be analysed either visually or through touch. Even though texture is an intuitive concept, an exact definition of texture either as a surface property or as an image property has never been adequately formulated (Rao 1990), (Heaps and Handel 1999). In the context of our research, texture is considered as a visual percept and is defined as *a visual pattern that exhibits high homogeneity.*

In vision research to date, there are two main computational approaches to the analysis of texture (Tomita and Tsuji 1990): the *statistical* approach and the *structural* approach. The statistical approach relies primarily on pre-attentive viewing (i.e. when textures are viewed in a quick glance) and is based on statistics and probability. In the structural approach, a texture is defined as being composed of a primitive pattern that is repeated periodically or quasi-periodically over some area. The relative positioning of the primitives in the pattern is determined by placement rules. Francos *et al* (1991) describe a texture model which unifies the statistical and structural approaches. Their model allows

the texture to be decomposed into three orthogonal components: a harmonic component, a global directionality component, and a purely non-deterministic component. An analogy might be drawn between this approach to visual texture and a sound synthesis technique proposed by Serra (1997b) based on a deterministic plus stochastic model.

Although texture has been studied extensively in various research areas, there is a small number of studies that attempted to identify relevant perceptual dimensions of visual texture. Early studies (e.g. Tamura *et al* (1978), Amadasun and King (1989)) focused on how to improve computer systems in order to match the human visual system in texture identification and classification experiments. To this end, these studies constructed various sets of texture dimensions based on the statistical analysis of textures (for example, analysis of grey-level co-occurrence matrices). The participants in these experiments were asked to rate a small number of textures along those dimensions and their ratings were compared with the ones obtained by computer-based texture identification and classification. Tamura *et al* (1978) suggest *coarseness, contrast,* and *directionality* as the most prominent perceptual dimensions of texture. Amadasun and King (1989) propose *coarseness, contrast, busyness, complexity,* and *texture strength* as appropriate texture dimensions that correspond to texture perception by humans. In a different approach, Ware and Knight (1992, 1995) propose a visualisation method based on a structural model of visual texture where each texture primitive is assumed to be a Gabor function. According to Ware and Knight, various mathematical parameters of Gabor functions can be used to describe dimensions of visual texture such as *orientation, size,* and *contrast.* However, this latter approach has not been empirically validated.

A limitation of the above studies is that the proposed dimensions were not empirically derived but based on the authors' subjective views. Thus, the question of whether humans use these dimensions in texture judgements was not adequately answered. In order to address this limitation, Rao and Lohse (1993, 1996) performed a series of experiments that tried to identify the high-level dimensions of texture perception by humans. Their studies were different to earlier studies of visual texture perception in that they used a variety of experimental designs and statistical methods (for example, multidimensional scaling techniques) that have been shown to provide an objective and appropriate way of investigating multidimensional phenomena. Rao and Lohse confirmed some of the dimensions proposed by earlier studies as being prominent in visual texture perception and suggested a strong correlation between certain sets of dimensions. The perceptual space proposed by Rao and Lohse (1996) comprises the following three dimensions:

- Repetitiveness. This dimension refers to the way primitive elements are placed and repeated over a texture image. The degree of repetition (e.g.

periodic, quasi-periodic, random) can be specified and controlled by placement rules.

- **Contrast and Directionality**. Contrast is related with the degree of local brightness variations between adjacent pixels in an image (i.e. sharp vs. diffuse edges). The directionality of a texture is a function of the dominant local orientation within each region of texture. These dimensions were found to be highly correlated in the experiments described in Rao and Loshe (1996).

- **Coarseness, Granularity, and Complexity**. Coarseness refers to the size (large vs. small) and density (coarse vs. fine) of the texture elements. Texture granularity corresponds to the degree of randomness with which the texture elements are distributed across the texture image. According to Rao and Lohse, "complexity is a harder feature to capture computationally" (p. 1667), and in their study they suggested the time taken for a subject to describe a texture and the suitability of the description as indicators of complexity. These dimensions are very similar to each other, a fact that is supported in Rao and Lohse's study by high correlation coefficients between these dimensions.

Based on the above we can suggest that the model proposed by Rao and Lohse (1996) satisfies our criterion of empirical support. An important characteristic of the above perceptual space is the identification of composite dimensions. The question of orthogonality is very important for the effective use of texture in visualisation. Although Rao and Lohse suggested a strong positive correlation between contrast and directionality, we approach this conclusion with some caution. As an example, consider the texture images depicted in Figure 5.1. Images (a) and (b) exhibit different levels of contrast but the same degree of orientation. We could suggest that the proposed correlation between contrast and directionality was a function of the texture images used in the above-described study. However, this is not the same for the composite dimension of coarseness, granularity, and complexity since correlation between these dimensions has been suggested in other studies as well (e.g. Tamura *et al* (1978)).

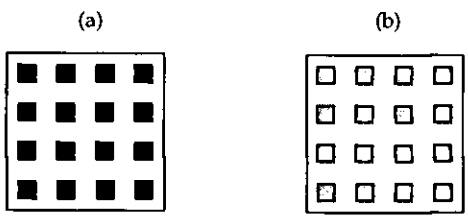


Figure 5.1: Images (a) and (b) have the same orientation but differ in their amount of contrast.

A question that arises is whether the above-described dimensions of visual texture are orthogonal to dimensions of colour. Although our perception of visual texture is to a large degree independent of colour, there is significant interaction between texture contrast and colour brightness since the former is primarily based on the brightness of the pixels as discussed above. This latter colour-texture interaction is discussed in more detail in the next chapter. As far as the criteria of measurability and synthesisability are concerned, although Rao and Lohse have not proposed measurement and synthesis methods for their perceptual space, various formulae have been suggested in the other studies discussed in this section. As a result of the above discussion we propose a refined set of texture dimensions that is based on Rao and Lohse (1996) and comprises *repetitiveness*, *contrast*, and the composite dimension of *coarseness*, and *granularity*. In our research, texture complexity is assumed to be a function of coarseness and granularity.

5.2 Visual Texture and Timbre

Timbre and visual texture have been shown to share some very important characteristics. First, timbre and visual texture are both multidimensional perceptual phenomena that can be described by a small set of prominent dimensions. Second, studies in both fields are based on a similar research methodology, i.e. rigorous empirical investigations of how humans perceive and describe sensory percepts. We evaluated the findings of these studies using a number of important criteria (empirical support, independence, measurability, and synthesisability). Table 5.1 summarises the two sets of perceptual dimensions we propose as suitable for further investigation.

Timbre	Texture
Sharpness	Repetitiveness
Compactness	Contrast
Sensory Dissonance	Coarseness-Granularity

Table 5.1: Perceptual dimension sets for timbre and visual texture.

The next step in our research was to design and conduct an experiment based on these sets of dimensions for timbre and visual texture. The objective of this experiment was the identification and investigation of associations (if any) between these auditory and visual dimensions.

5.2.1 Method

Experimental Design

We used a within-subjects experimental design, where each participant performed a series of texture-timbre association tasks for all the perceptual dimensions of timbre and visual texture. The order of the tasks was random for each subject in order to control possible ordering effects.

Subjects

As discussed in the previous chapter, a limitation of our first empirical investigation was the assumption that musical experience will have no effect on subjects' responses. However, the results of that experiment indicated that there was a noticeable difference between the responses of music and non-music subjects. Therefore, at this research stage we decided to conduct the texture-timbre experiment with one subject group consisting of subjects with a strong musical background.

We had twenty subjects in total and all were given a screening questionnaire about their experience in both traditional and computer music (see Appendix A for a sample copy of the questionnaire). The exact composition of the twenty subjects was:

- Sixteen undergraduate students at Middlesex University's Sonic Arts department.
- Three lecturers at Middlesex University's Sonic Arts department.
- One professional composer.

Although four subjects had previously taken part in the experiment described in the previous chapter, we did not anticipate any *carry-over* effect on those subjects' responses since timbre and visual texture were not part of that experiment. Furthermore, in the texture-timbre experiment, colour information was excluded and the sound stimuli were approximately equalised for pitch and loudness as discussed later.

Apparatus and Stimuli

A prototype computer application was designed in Macromedia Director v. 7.0 (Macromedia 1998) for use in this experiment comprising a texture palette and three series of sound sequences.

The texture palette consisted of three series of texture sequences — one for each perceptual dimension. There were three variation modes (ascending, descending, non-monotonic) for each series, thus giving nine sequences in total. Each texture sequence consisted of five texture images. The individual texture images were based on a simple texture pattern that was then modified to represent the variation in each perceptual dimension (see Figure 5.2).

Contrast manipulation is a standard feature in most image-processing software tools and for this experiment we used Adobe Photoshop v. 4.0 (Adobe Systems 1996) as the primary design tool (Figure 5.2 - top). In addition, we controlled the degree of texture repetitiveness by increasing or decreasing the number of displaced texture elements from an ideal symmetrical grid as shown in Figure 5.2 - centre. Finally, the combination of gradually smaller texture elements with increasing levels of density and visual noise provided the control for the composite dimension of coarseness and granularity (Figure 5.2 - bottom). Note that the texture sequences used in this experiment represented variation in a single dimension, i.e. one dimension was varying whilst the remaining two were kept constant.

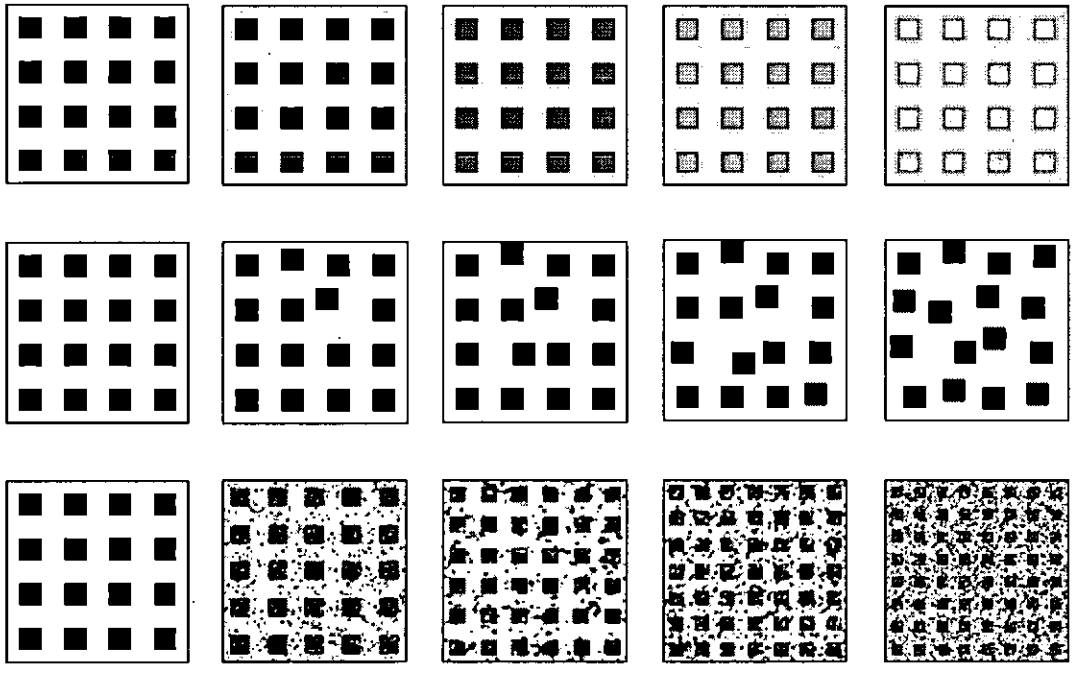


Figure 5.2: The texture images used in this experiment: (Top) Contrast, (Centre) Repetitiveness, (Bottom) Coarseness-Granularity.

Three series of sound sequences were designed — one for each perceptual dimension of timbre. The same variation modes were used as for the design of texture sequences and each sequence consisted of five sounds. Sharpness and dissonance sequences were designed using *TurboSynth* (Digidesign 1985) while for the design of compactness sequences we used *Csound* (Vercoe 1993). All sounds had a fundamental frequency of 220

Hz and the same amplitude level for pitch and loudness equalisation respectively. The duration of each sound was one second and sounds within a sequence were separated by 0.5 seconds of silence. Sharpness sounds were designed through the gradual addition of harmonic frequency components up to the sixth in the harmonic series (i.e. the highest frequency component was: 220 Hz × 6 = 1320Hz). Dissonant sounds consisted of six frequency components deviating from the harmonic series in order to produce beating among adjacent partials. The degree of beating was measured with the formula described in Hutchinson and Knophoff (1978). In order to produce the *tone-noise* effect for the dimension of compactness we used noise bands centred on the above six harmonic frequency components and gradually increased the noise bandwidth.

Experimental Task

The experimenter demonstrated how to use the prototype application (see Figure 5.3). This was followed by a practice period of three sound sequences. The experimental task was: for the current sequence of five sounds to create a sequence of five corresponding textures. Subjects could listen to the current sequence as many times as they wished, at any point during the task. During the experiment, both sound and texture sequences were introduced in a different order for each subject (a random number generator was used to create random sequence orders). Each subject completed the task for nine sequences. Subjects performed the experiment at their own pace and times ranged from twenty to thirty minutes. The experimenter was present throughout the experiment recording observations that formed the basis for post experiment interviews with subjects. Finally, a data collection program logged texture selections in the form of screenshots, as well as completion time per sequence.

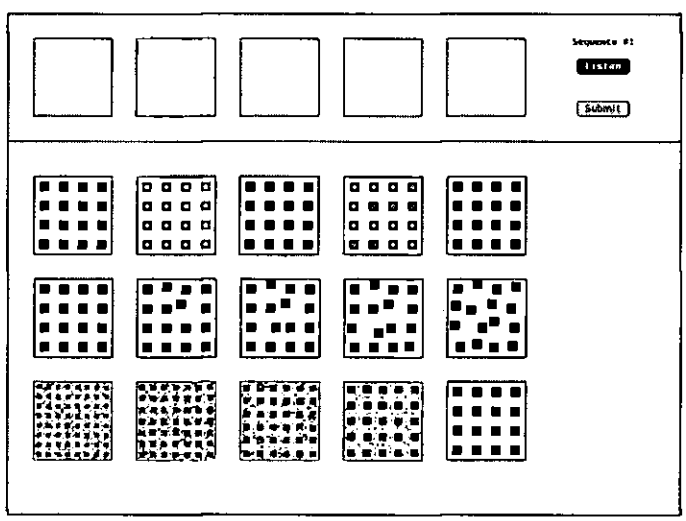


Figure 5.3: The prototype application used in this experiment. Texture sequences were arranged horizontally. Subjects could drag images onto the empty display area (see top of image).

Experimental Environment

The experiment was conducted in a room with normal 'office' lighting and sounds were presented binaurally through headphones. The experiment was designed and run on an Apple Power Macintosh G3 personal computer. Subjects sat approximately 80cm away from the computer screen and the components of the interface were sized for comfortable viewing and manipulation at that distance.

5.2.2 Analysis of Results

As in the case of our first empirical investigation with colour, the presentation of the results for the above-described experiment is based on a qualitative method supported by quantitative data. The major qualitative variable is the texture selection strategy followed by subjects. Table 5.2 shows overall results for each dimension of timbre after the processing of the raw data obtained from subjects' texture selections independently of the three variation modes. For all three dimensions of timbre, the obtained results indicate strong associations between selection strategies that involved variation along a *single* dimension of visual texture as shown in the first three rows of results in Table 5.2. Note that, for simplicity of presentation, the composite dimension of texture coarseness and granularity has been abbreviated as CG. Figures 5.4 - 5.6 are based on the texture sequences created by subjects and display average levels for each stimulus within a texture sequence for each of the dimensions of timbre and variation modes. These figures should be interpreted as showing (a) how close were the average subjects' responses to the desired variation levels, and (b) an indication of the correlation (positive, negative, none) between subjects' sequences and the sequence stimuli.

	Timbre Dimensions		
Selection Strategy	**Sharpness**	**Compactness**	**S. Dissonance**
Contrast	80	1.67	11.67
CG	3.33	95	6.67
Repetitiveness	11.67	0	80
Mixed	5	3.33	1.67
None	0	0	0
Total (%)	100	100	100

Table 5.2: Overall results for sequences varying in any of the dimensions of timbre.

In more detail, there is a clear preponderance of *sharpness-contrast* associations (80% of all sequences) as opposed to associations between sharpness and other texture dimensions. Figure 5.4 shows the average contrast sequences created by subjects as a response to each sharpness variation mode. A very strong positive correlation between the variations in sharpness and contrast can be noticed in all variation modes. Based on the above results we can suggest that for the majority of subjects, textures with low levels of contrast were associated with dull sounds while higher levels of contrast were associated with sharper sounds.

In the case of sound sequences with varying levels of compactness the dominant association strategy was to vary texture CG (95% of all sequences) as opposed to a total 5% of associations between compactness and other texture dimensions. Figure 5.5 shows the average CG sequences created by subjects as a response to each compactness sequence. These results suggest a very strong positive correlation between compactness and texture CG in all varying modes. Overall, coarse textures with large elements were associated with tone-like sounds while finer textures with small elements were associated with noise-like sounds.

Similarly, in the case of sound sequences with varying sensory dissonance, the results suggest a strong *sensory dissonance-texture repetitiveness* correspondence (80% of all sequences). Figure 5.6 shows the average repetitiveness sequences created by subjects as a response to each sensory dissonance sequence. A strong positive correlation between sensory dissonance and texture repetitiveness can be noticed in all variation modes. Regular textures were associated with sounds composed of harmonically related partials while irregular textures corresponded to inharmonic sounds with increasing *beating* among adjacent partials.

In addition, Figure 5.7 shows maximum, minimum, and mean response times for each dimension of timbre and each variation mode. The mean response times provide further evidence for the *intuitiveness* of subjects' responses. Average times were fast (<60 seconds) for all three dimensions of timbre and variation modes although subjects were slower when presented with non-monotonic sound sequences. These results led us to conclude that subjects were at ease with the experimental task and responded instantly without having to listen many times to the sound stimuli.

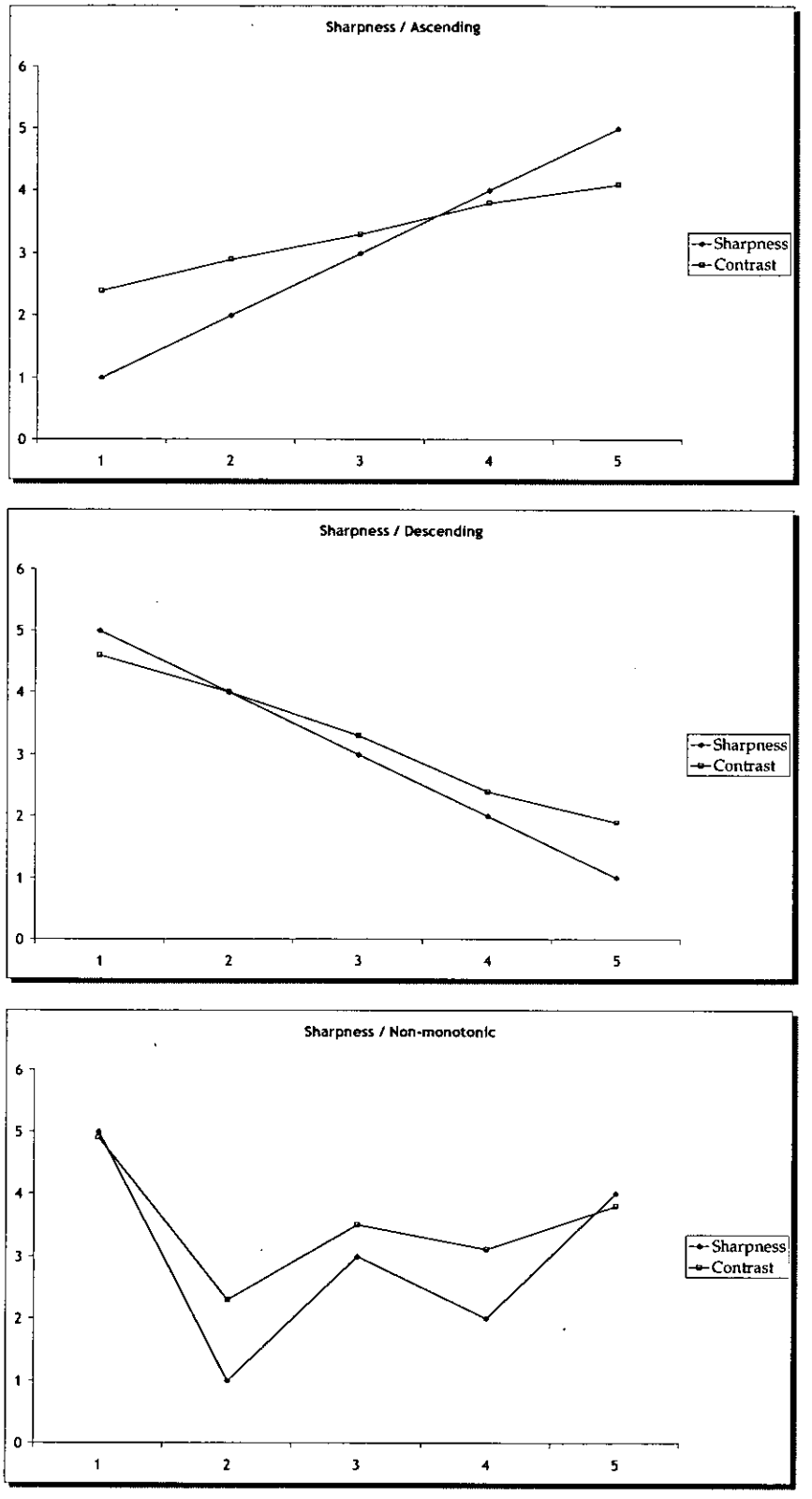
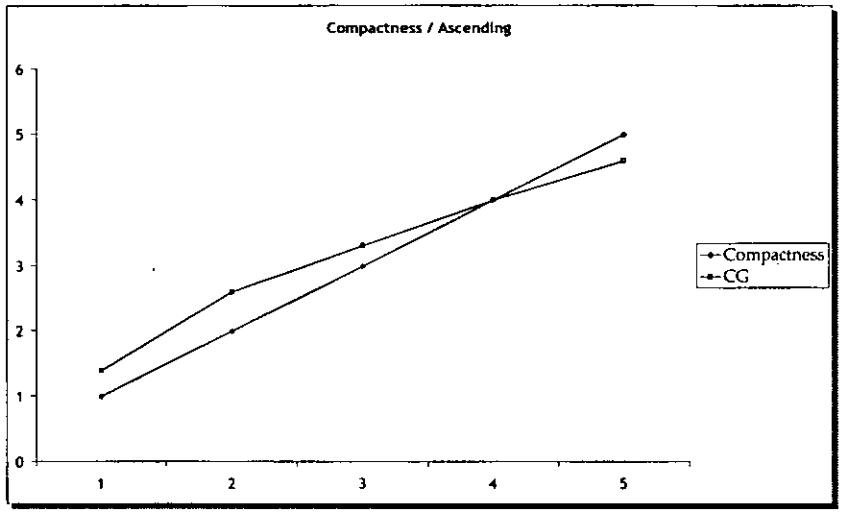
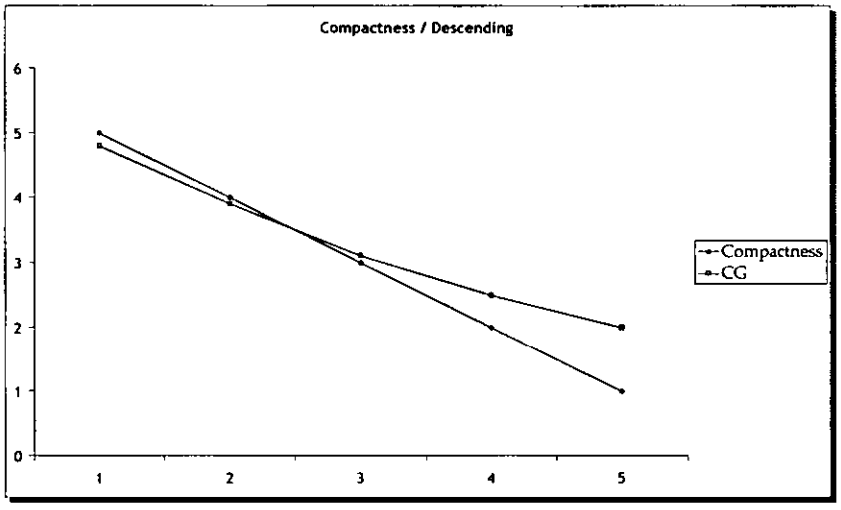


Figure 5.4: Average subjects' responses in texture contrast for sharpness stimuli (all variation modes).





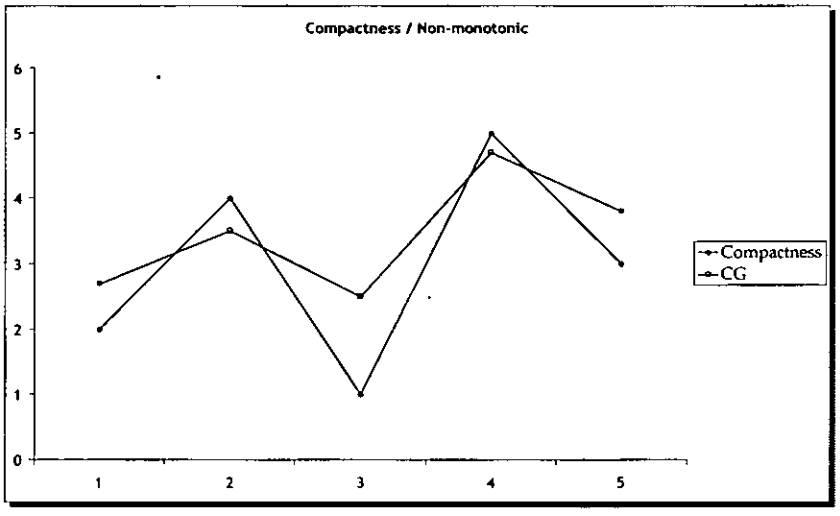


Figure 5.5: Average subjects' responses in texture CG for compactness stimuli (all variation modes).

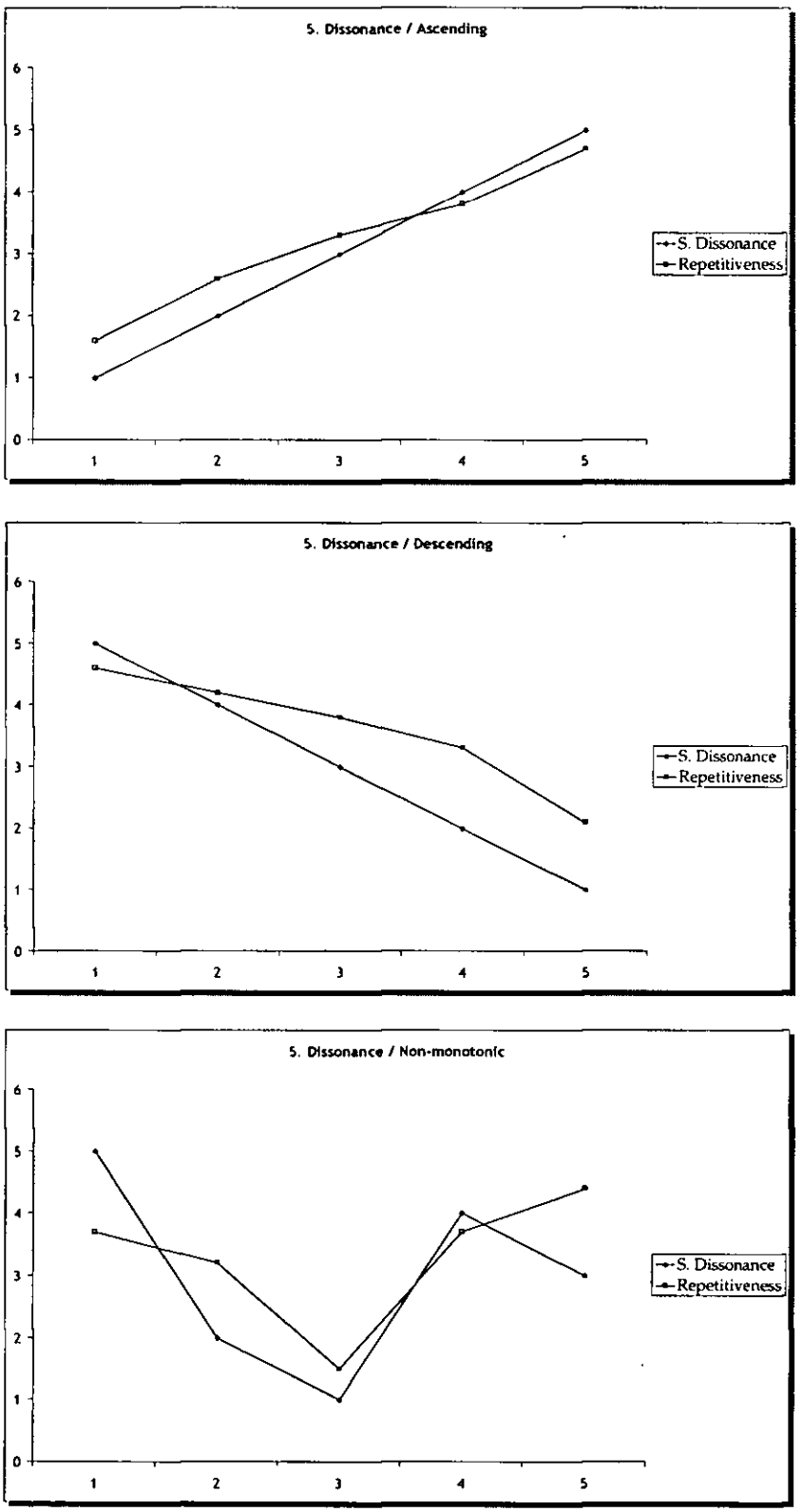


Figure 5.6: Average subjects' responses in texture repetitiveness for sensory dissonance stimuli (all variation modes).

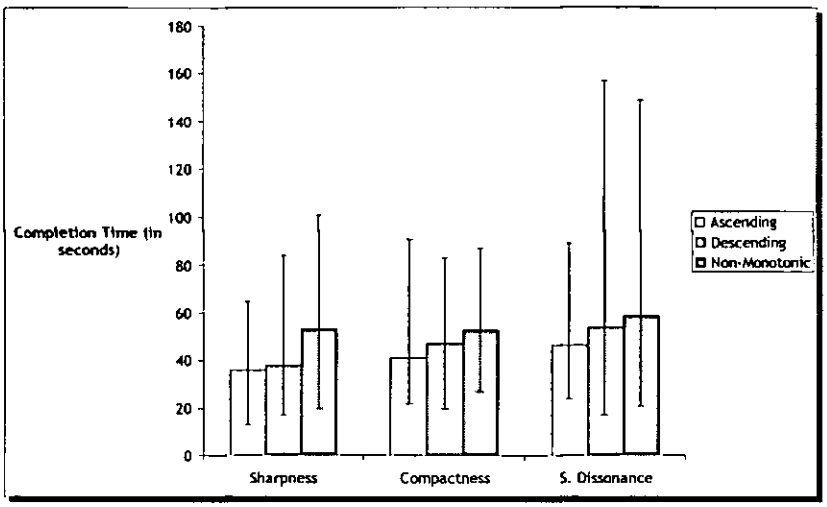


Figure 5.7: Maximum, minimum and mean completion times for each dimension of timbre according to each variation mode.

5.3 Conclusion

Based on the above analysis and discussion we can construct a three-dimensional space for the associations between timbre and visual texture as depicted in Figure 5.8. In this space, the dimension of sharpness is associated with variations in texture contrast, sensory dissonance corresponds to texture repetitiveness, and compactness is associated with the composite dimension of texture coarseness and granularity.

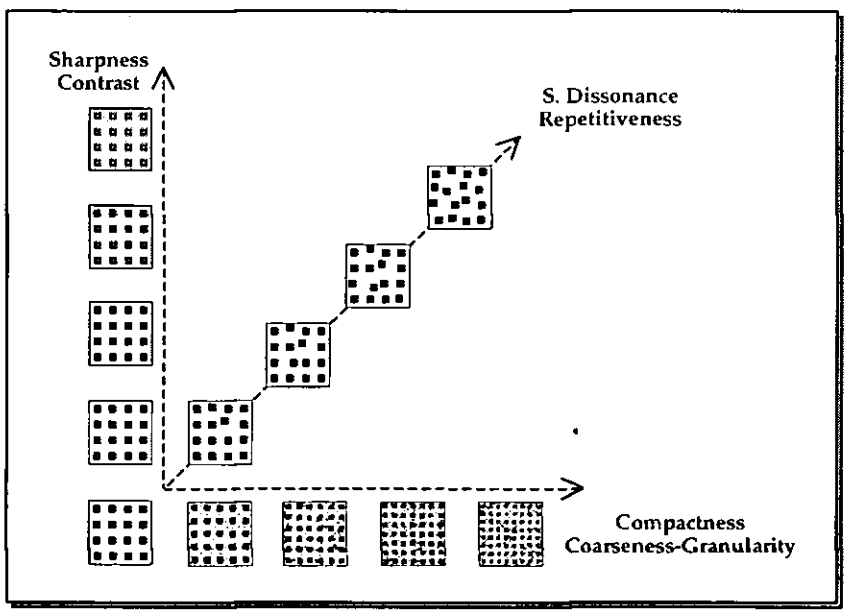


Figure 5.8: The 3-D space proposed for the associations between perceptual dimensions of timbre and visual texture.

A limitation of the empirical investigation described in this chapter arises from the small dimension sets used in the investigation. Ideally, a high number of perceptual dimensions are required for the complete description of both timbre and visual texture. Therefore, the obtained results are limited to the dimensions used in our experiment. In addition, mention should be made that the obtained results can be only suggestive due to the small number of subjects that participated in our experiment. Finally, it is of further interest to investigate the validity of the above texture-timbre associations with non-music subjects, an issue that is addressed in later chapters of this thesis.

6

Sound Mosaics I

This chapter is organised into two main sections. In the first section, we propose a novel framework for sound visualisation that is formulated upon the findings of our empirical investigations as described in the previous two chapters. Our visualisation framework is embodied in *Sound Mosaics*, a prototype user interface for sound synthesis based on direct manipulation of visual representations. The details of an initial implementation of Sound Mosaics are described in the second section of this chapter.

6.1 A Novel Framework for Sound Visualisation

This thesis presents a novel method for using dimensions of visual perception to visualise auditory percepts. Studies in the research areas of auditory and visual perception have identified a small number of dimensions that are considered important for the perception of visual and auditory information. A first step in our research was to define sets of prominent perceptual dimensions for both sound and image. In Chapter 3 we derived a model of auditory perception comprising the dimensions of *pitch*, *loudness*, and *timbre*. In this model, pitch and loudness were considered to be uni-dimensional whereas timbre as a multidimensional phenomenon comprised the steady-state auditory dimensions of *sharpness*, *compactness*, and *sensory dissonance*. In a similar manner, our review of visual perception studies (see §4.1 and §5.1) assisted us in the formation of a model of visual perception comprising *colour* and *visual texture*. Since both colour and texture are multidimensional phenomena, we further constructed two sub-dimension sets for these visual percepts. Colour was described in terms of three perceptual attributes, namely *hue*, *saturation*, and *brightness*. Visual texture was described in terms of *contrast*, *repetitiveness*, and the composite dimension of *coarseness* and *granularity* (for simplicity of presentation, this latter dimension of visual texture is abbreviated as CG). Furthermore, we conducted a series of controlled experiments to investigate associations between the above sets of auditory and visual dimensions as described in the previous two chapters. The results of our empirical studies suggested a number of important auditory-visual associations as illustrated in Figure 6.1.

Our method uses the colour dimensions of brightness and saturation to visualise auditory pitch and loudness respectively. In more detail, participants in our experiments associated dark colours with low-pitched sounds and light colours with high-pitched sounds. Weak colours were associated with soft sounds whereas louder sounds corresponded to stronger or more vivid colours. Our empirical results also presented evidence that in both cases colour hue was kept constant and therefore at this stage of our research we cannot suggest any associations between hue and the auditory dimensions of pitch and loudness.

Furthermore, we use visual texture to visualise perceptual dimensions of timbre. Our empirical results indicated that texture contrast was associated with auditory sharpness. Sharp sounds were associated with high-contrast textures whereas low-contrast textures were associated with dull sounds. In addition, texture repetitiveness was associated with sensory dissonance. In this case, regular textures corresponded to sounds with harmonically related frequency partials, whereas increasing deviation from the harmonic series was associated with textures increasing in their irregularity. Finally, the composite dimension of CG was associated with timbre compactness. Tone-like sounds were associated with non-granular simple textures composed of large elements, whereas granular and more complex textures with small elements corresponded to noise-like sounds.

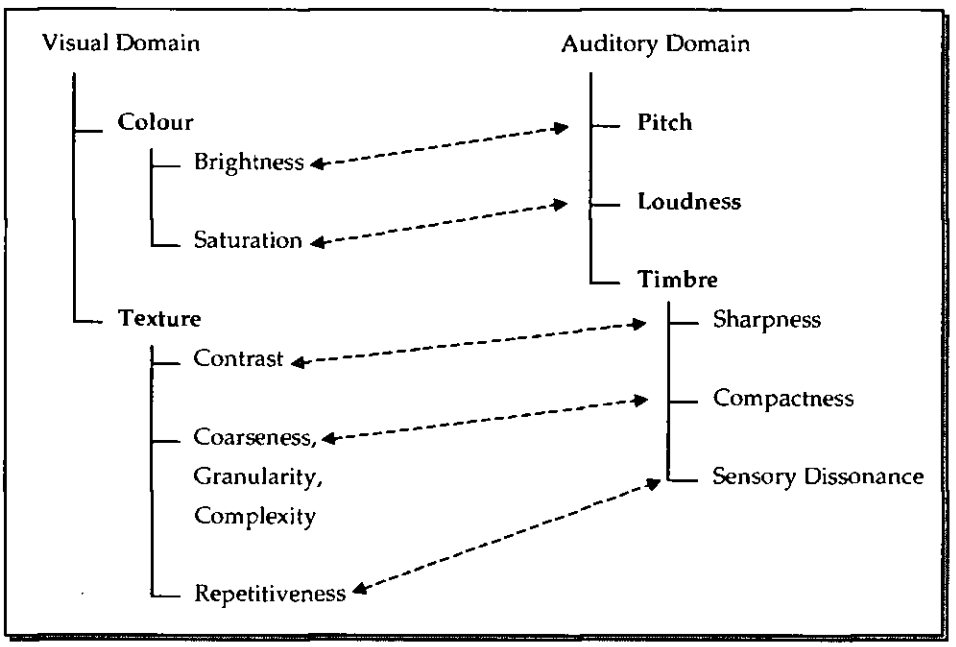


Figure 6.1: The proposed framework for associations between auditory and visual percepts.

The above-discussed auditory-visual associations formed the theoretical framework for the design and implementation of a novel user interface for computer-based sound synthesis as described in the next section.

6.2 Initial Implementation of Sound Mosaics

6.2.1 Overview of the Implementation

Sound Mosaics has been implemented in Java using the Sun Java Development Kit (JDK) 1.1.7B on an Apple Power Macintosh G3 personal computer with a 17" monitor capable of representing millions of colours at a resolution of 1024×768 pixels. As described in

more detail later in this chapter, the current implementation of Sound Mosaics incorporates Csound (Vercoe 1993) as its underlying sound processing engine. Csound is a widely used music programming language that is freely available and runs on the most popular operating systems. As a result of the above implementation decisions, Sound Mosaics can be designed to work on a variety of platforms with only a minimum amount of modification.

The initial Sound Mosaics prototype is illustrated in Figure 6.2. The application window is split into two main regions that support various actions.

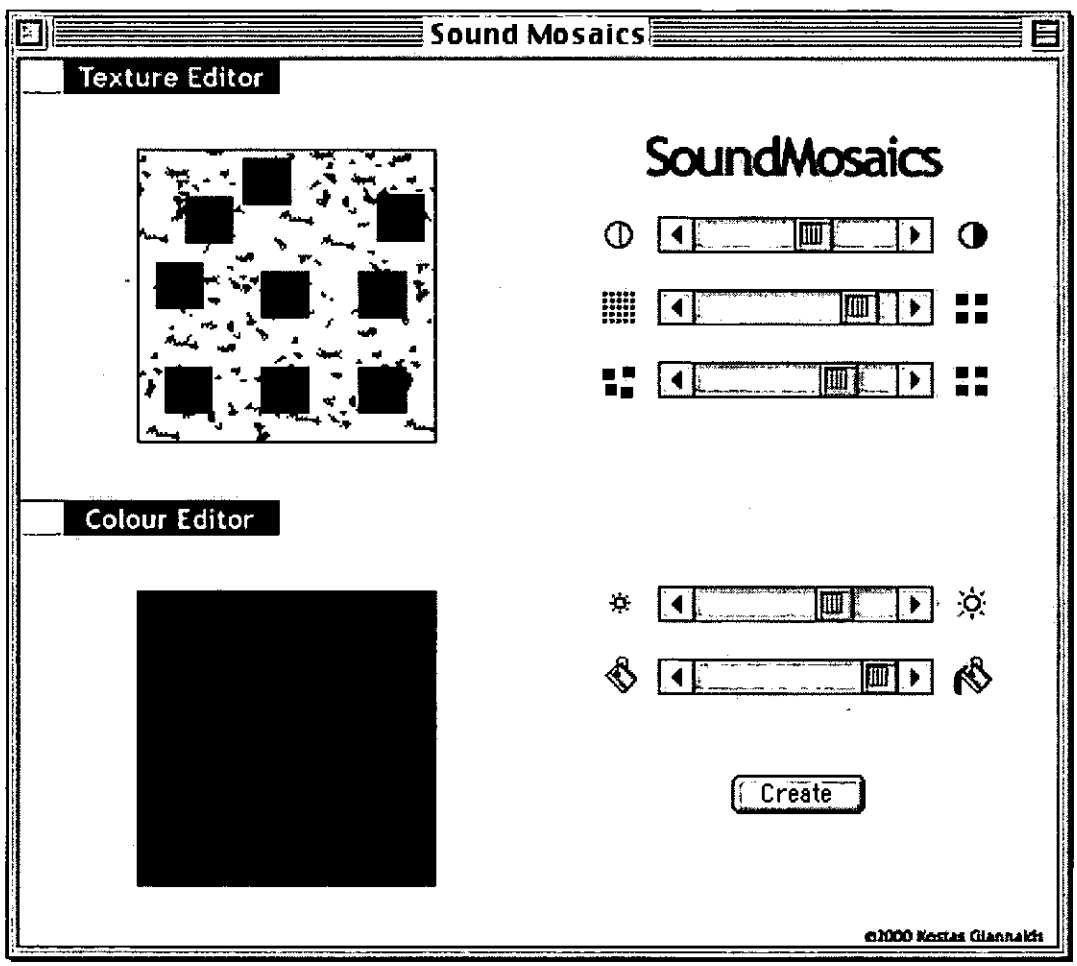


Figure 6.2: The initial Sound Mosaics prototype (see body of text for a full description and Colour Plate E.2).

The upper-half region contains the *texture editor* panel that displays a texture image (upper-left quartile) and includes three scrollbars (upper-right quartile) for the manipulation of the three perceptual dimensions of visual texture that are supported in Sound Mosaics. The top scrollbar adjusts the amount of contrast, the middle scrollbar

adjusts the level of CG, and finally the bottom scrollbar adjusts the level of texture repetitiveness. In a similar manner, the lower-half region of the application window contains the *colour editor* panel that displays a colour image and includes two scrollbars (lower-right quartile) for the manipulation of colour dimensions. The top scrollbar adjusts the amount of colour brightness, and the bottom scrollbar adjusts the level of colour saturation. Both texture and colour images can be modified only through the use of scrollbars. Special icons have been designed to describe the functionality of each scrollbar in a visual manner. Textual guidance has been completely excluded in order to measure the extent to which users can extract information about the auditory parameters from purely visual representations. The background colour of the application window has been set to an aesthetically pleasing colour (orange) which appears not to affect the perception of the texture and colour images. Note should be made that the design choice of scrollbars in Sound Mosaics was made to avoid implementation issues that arise from allowing users to interact directly with the texture and colour images. User actions such as drawing would require a sophisticated image processing stage that was outside the current scope of our research and has been left for further work (see §9.5). Although, modifications on the texture and colour images take place in real time, at present the processing of the auditory information cannot be processed in a real-time manner. Instead, users have to click on the "Create" button (lower-right quartile) in order for Sound Mosaics to map the current visual parameters to their auditory counterparts and pass the information to Csound which plays the corresponding sound immediately after it has been synthesised. The time between users clicking the "Create" button and hearing a sound may be up to two seconds (on an Apple Power Macintosh G3 running at 300MHz) depending primarily on the complexity of the synthesis parameters.

In this implementation of Sound Mosaics, a sound object has a dual visual representation, one for the representation of timbre and the other for the representation of pitch and loudness. Although this design idea seems counter-intuitive and a single representation is preferable to a dual one, our design choice was influenced by the interaction between texture contrast and colour brightness that occurs when the two representations are combined into a single one. As an example, if we were to represent both colour and texture information on the same image, the background layer could have been used to display the colour information whereas texture would be drawn on the foreground layer. Each time the user wished to adjust colour brightness (or auditory pitch) the change would affect the perceived texture contrast (or auditory sharpness) thus violating the independent control of these visual dimensions. In order to avoid this problem, this implementation of Sound Mosaics displays visual representations of sound that consist of two separate images. Nevertheless, since timbre, pitch and loudness are considered to be independent, it can be argued that the dual representation approach may be more suitable in a familiar situation where the music composition process is split into two parts:

- First, the design of an 'instrument' based on the desired timbre characteristics.

- Second, the choice of a desired pitch and loudness level for the instrument to be played on.

The core of the Sound Mosaics implementation comprises two components for image and sound synthesis as described in the following sections.

6.2.2 The Image Synthesis Component

In Sound Mosaics the image synthesis component is built around two mechanisms for the generation of visual texture and colour information.

The Colour Mechanism

The colour mechanism has been designed to manipulate the two perceptual dimensions of colour supported in this implementation of Sound Mosaics, namely brightness and saturation. Manipulation of colour parameters is common and straightforward in programming environments such as the JDK. In particular, JDK 1.1.7B provides explicit control over hue, saturation, and brightness in floating-point values ranging from 0.0 - 1.0 (or 0% - 100%). The underlying model is based on the HSB colour space as described in §4.1. Our own informal experimentation with different levels along the brightness scale suggested that values between 0.0 - 0.3 were very hard to discriminate and therefore we chose the value of 0.3 as the lowest threshold for brightness. The range between 0.3 - 1.0 was split into 70 equal steps in order to cover the large range of pitch information that should be available to users. Saturation changes were more perceivable throughout the scale and we have therefore used the whole scale (100 equal steps). In addition, we have used a constant hue because the results of our first experiment did not provide a sound basis for associations between hue and any auditory parameters. Since human colour discrimination is better in the *red-green* region than other regions of the visible light spectrum (Fortner and Meyer 1997) we have chosen to use a constant red hue.

One limitation of our approach is related to the number of just noticeable differences that we can perceive in any of the above colour dimensions. Although we can perceive a very large number of colours and current computer displays are designed to reproduce millions of colours, colour discrimination varies significantly with the viewing conditions. For example, if two colour patches are simultaneously presented we can detect very slight differences in any of the perceptual dimensions of colour. However, if the two patches are observed in isolation, then a much larger difference is needed in order to perceive the two colours as different. Some sources state that the number of

colours to be used for visualisation purposes should be as small as possible and give approximate ranges for each perceptual dimension. In the initial implementation of Sound Mosaics, we chose to use as many values as needed to represent an adequate range of auditory information.

Another limitation of our approach arises from the fact that the HSB colour space is not perceptually uniform (Fortner and Meyer 1997). Equal steps in any of the three dimensions do not necessarily correspond to equally perceived changes. Although perceptually uniform colour spaces do exist (for example the CIE LUV space proposed by the Commission Internationale de L'Èclairage (CIE) in 1976) these are harder to implement on computers (Foley *et al* 1994) and we chose to investigate them in future research. We did, however, explicitly attempt to measure the extent of the above limitations in our evaluation of Sound Mosaics (described later).

The Visual Texture Mechanism

The visual texture mechanism manipulates the three perceptual dimensions of texture supported in Sound Mosaics: contrast, repetitiveness, and the composite dimension of coarseness and granularity. Although various texture generation algorithms have been suggested in the related literature, they are primarily based on mathematical models (or other) rather than on explicit control of perceptual dimensions. These models are capable of generating very realistic images resembling the properties of natural textures. A problem associated with this is the existence of image characteristics that were not investigated in our study (e.g. depth effects, 3D, etc.) and therefore we have no control over them. For these reasons, we chose to implement a set of algorithms that generate textures in terms of the three perceptual dimensions of texture included in our research. The resulting textures belong to a class of textures called *abstract* (or *synthetic*) textures due to their non-resemblance to natural textures. Synthetic textures can represent perceptual dimensions of natural textures in a controlled manner and they have been used extensively in visual texture perception studies (Heaps and Handel 1999).

In Sound Mosaics, the texture image is divided in two layers as shown in Figure 6.3. The first layer is a background area upon which the second layer (foreground) consisting of a number of texture elements is drawn. Both background and foreground layers are achromatic, i.e. they range between black and white. The overall image size has been chosen to be a 128×128 image that allows future implementations to take advantage of common image-processing tasks (for example, Fourier analysis requires powers of 2).

The contrast mechanism works by adjusting the brightness levels of the background and foreground layers on the texture image. As described in §5.1, contrast is related to the degree of local brightness variations between adjacent pixels in an image. Therefore,

when the brightness variation between the background and foreground layers is small (approximately halfway through the black-white scale), the contrast level is low and the elements' edges are diffused in the background layer (Figure 6.3 - left). Maximal brightness variation occurs when the background and foreground layers are painted white and black respectively (Figure 6.3 - right). The above approach is a simplified contrast control as most natural textures are usually composed of a large number of different brightness levels.

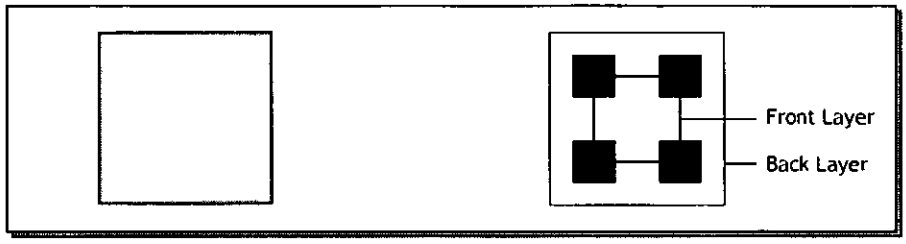


Figure 6.3: A texture image in the initial Sound Mosaics prototype consisted of two layers as shown in the right part of the figure. The figure also shows the two extremes of the texture contrast scale.

The mechanism to control the composite dimension of CG works in the following way. We have used the number and size of the texture elements as the main indicators of texture coarseness whereas the presence of random elements (visual noise) is used to increase the perceived granularity and complexity of the texture image. Coarse textures consist of a small number of elements that are large in size (Figure 6.4. - left). Increasing the number of texture elements while decreasing their size and introducing visual noise leads to successively finer and more complex textures (Figure 6.4 - right).

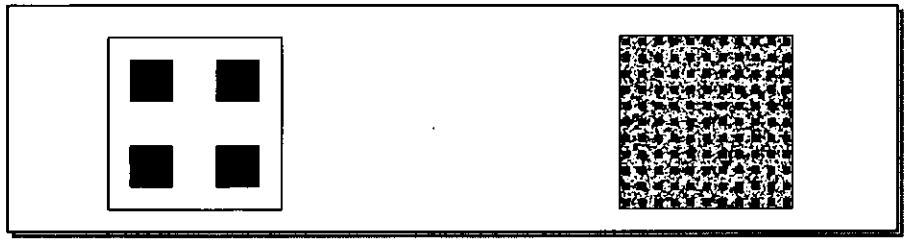


Figure 6.4: The two extremes of the texture CG scale. Note that the depicted images have maximal contrast and repetitiveness.

Texture repetitiveness refers to the structure of the texture elements. In Sound Mosaics, placing the elements on a symmetrical grid generates regularly structured texture elements (Figure 6.5 - left). Conversely, displacing the elements from their anchor points increases texture irregularity (Figure 6.5 - right). In more detail, Sound Mosaics selects a

number of texture elements at random according to the desired level of repetitiveness and displaces those elements by a random distance at a random direction within the confines of the overall image size. The number of elements to be displaced increases with the level of irregularity so at maximal irregularity, all elements are being displaced. A similar approach is discussed in Healey and Enns (1998).

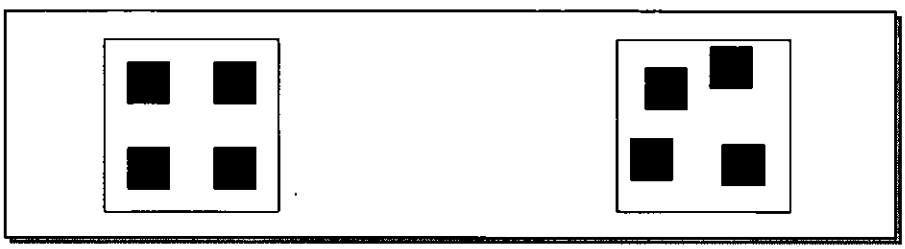


Figure 6.5: The two extremes of the texture repetitiveness scale. Note that in the depicted images have maximal contrast and CG.

As in the case of colour dimensions, a limitation of our texture generation mechanisms is that the number of just noticeable differences is not known so our approach is prone to errors. Furthermore, the shape of the texture elements plays an important role in texture perception and human vision in general. However, since our research has not investigated perceptual dimensions of shape we chose to use a constant familiar shape, in this case, square texture elements.

6.2.3 The Sound Synthesis Component

As discussed earlier in §3.1.3, timbre has been shown to depend on certain characteristics of the sound spectra. Based on this we can suggest that spectral models for sound synthesis can form the basis for a more intuitive approach to sound design. Spectral models are based on perceptual reality, can be controlled by perceptual parameters, and they provide a general model for all sounds in an analysis/synthesis form (Serra 1997a,b). However, research in this area is still young and there is no definitive set of appropriate perceptual parameters that can be used effectively in sound synthesis systems.

One of the most powerful spectral modelling techniques is *additive synthesis* (Roads 1996; Miranda 1998). Additive synthesis refers to a class of techniques that perform sound synthesis by adding elementary waveforms to create a more complex waveform. Figure 6.6 illustrates the design of a general model for additive synthesis based on a number of sinewave oscillators, each with separate frequency and amplitude functions. The output of each oscillator is added together to create the final output. An advantage of additive synthesis is that it provides complete, independent control over the behaviour of each

spectral component and given enough oscillators theoretically any sound can be synthesised (Dodge and Jerse 1997). However, this means that additive synthesis can be very demanding in computational resources (Roads 1996).

In Sound Mosaics, the sound synthesis component is based on a simple model of additive synthesis implemented in Csound. A Csound program consists of two text-based files. The first file is called the *orchestra* file and it contains a number of instruments that are defined by interconnecting various *opcodes* that refer to a large number of system-defined or user-defined sound synthesis/analysis modules. The second file is called the *score* file and it controls how the instruments specified in the orchestra file are to be played by passing the necessary parameter values (e.g. frequency and amplitude values) to the opcodes. The orchestra and sample score files for Sound Mosaics can be found in Appendix D. Sound Mosaics is designed to continuously update its score file in order to reflect any changes that occur from user interaction with the system.

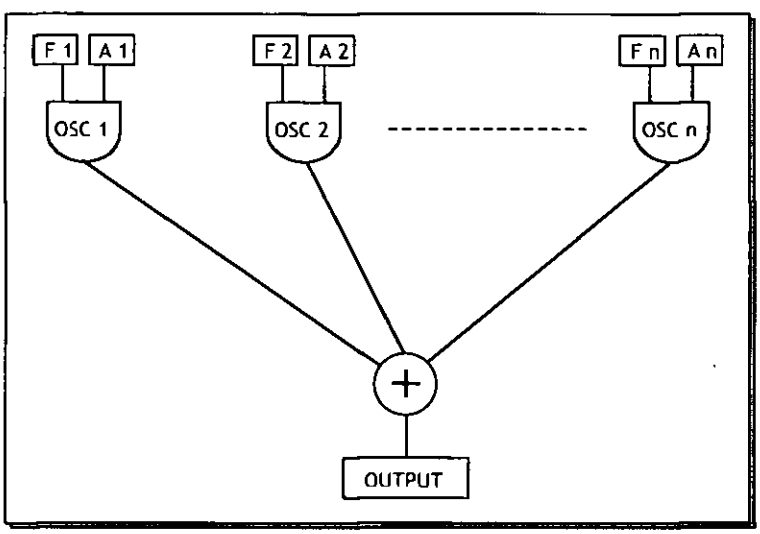


Figure 6.6: A general model of additive synthesis. The output from each of *n* oscillators, each with separate frequency (F) and amplitude (A) control functions is added together to produce the final output.

The additive synthesis model used in the initial Sound Mosaics prototype is capable of producing spectral components that do not change in their frequency or amplitude during the evolution of a sound object, an approach that is similar to what is known as *fixed-waveform* synthesis (Roads 1996). Although this is a limiting factor for the kinds of sounds that can be produced with Sound Mosaics, our design choice is based on the steady-state model of timbre incorporated in our research. Furthermore, all spectral components have equal amplitudes, since amplitude variations have been left out for

future investigations. Finally, sounds produced with Sound Mosaics have a fixed duration of 1.5 seconds and the overall amplitude envelope depicted in Figure 6.7.

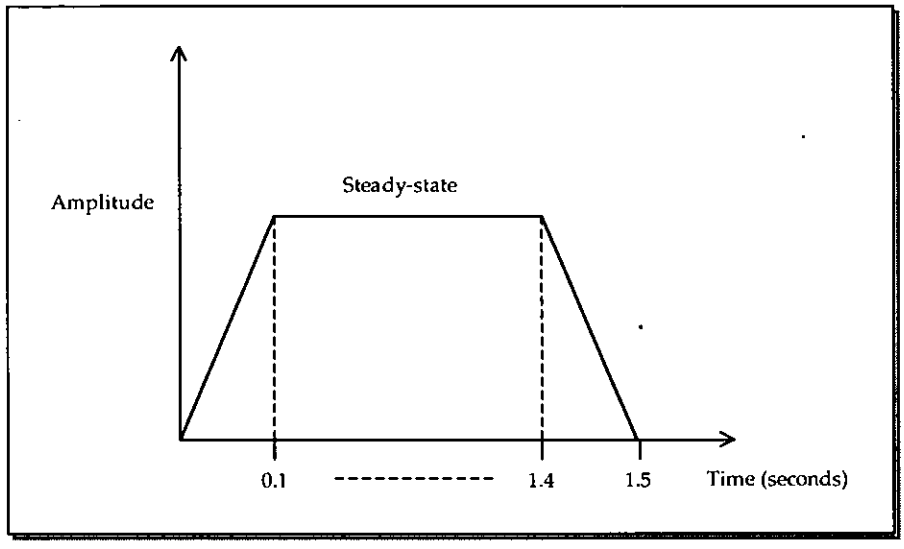


Figure 6.7: The overall amplitude envelope used for sounds created with Sound Mosaics. The attack and decay parts have been set to 0.1 seconds in order to eliminate the abrupt start and end of the sound.

We have designed synthesis algorithms for each of the auditory dimensions incorporated in the initial Sound Mosaics prototype as discussed in the remainder of this section.

The Pitch and Loudness Mechanisms

In Sound Mosaics, the pitch of a sound is solely determined by the value of the fundamental frequency component in the sound's spectrum. The range of pitch information that is represented in Sound Mosaics covers approximately six octaves. The lowest and highest pitches that can be produced with Sound Mosaics are 32.708Hz (or C_1) and 1864.655Hz (or $A\#_6$) respectively.

Although, the human hearing mechanism can detect frequencies in the range 20 Hz - 20000 Hz, pitch discrimination varies for different frequency ranges and some sources state that the overall number of just noticeable differences in pitch perception is 1400 (Olson 1967). The hearing range can be also expressed in octaves in which case it covers ten octaves. However, pitch discrimination becomes very difficult above 10000 Hz and the maximum accuracy for pitch labelling is about six octaves extending upward from about 60 Hz (Butler 1992). Therefore, our design choice of six octaves seems a reasonable one and resembles the pitch range of some music instruments (e.g. the piano covers

approximately eight octaves). A limitation of our approach is that it is bound to the Western music tradition.

Loudness is approximated through sound intensity as discussed earlier in §3.1.2. The lower and upper limits for sound intensity are 0 dB and 120 dB respectively. However, the use of the decibel scale is device dependent. In other words, the loudest sound that can be produced depends on the overall sound output level on the user's sound listening system. Therefore, the decibel scale is used here in the relative sense and not in absolute terms. The Csound intensity range (0 - 90dB) was split in 100 equal steps corresponding to the same number of saturation values. The previously discussed limitations associated with humans' perceptual discrimination also apply in the case of auditory dimensions.

The Timbre Mechanism

Sound Mosaics incorporates the model of timbre perception described in Chapter 3. This model treats timbre as a quality of the steady-state portion of sound excluding other envelope characteristics (e.g. attack and decay). The model comprises the dimensions of sharpness, compactness, and sensory dissonance.

As discussed in §3.1.3, the sharpness of complex tones, is determined by the upper limiting frequency and the way energy is distributed over the frequency spectrum, i.e. the higher the frequency location of the spectral envelope centroid, the greater the sharpness (Bismarck 1974a,b). Therefore, in the case of complex tones there are two ways that the sharpness of a tone can be specified and modified. First, if we assume that the tone partials remain constant in amplitude, then sharpness might be increased or decreased by adding or removing high frequency partials respectively. Second, if we assume that the number of partials is fixed, then the amplitudes of the partials can be modified in order to increase or decrease the perceived sharpness. As previously mentioned, amplitude variations have been excluded from our studies and therefore in the initial Sound Mosaics prototype, sharpness is controlled by the addition or removal of high frequency partials. Our sharpness synthesis algorithm is based on a spectral centroid (C) measurement formula proposed in Kendall and Carterette (1997):

$$C = \frac{\sum_{n=1}^{i} F_n A_n}{F_1 \sum_{n=1}^{i} A_n}$$

F_n: Frequency of n_{th} partial
A_n: Amplitude of n_{th} partial
F_1: Fundamental frequency

The sharpness algorithm computes the harmonic spectrum for the fundamental frequency given by the level of colour brightness. The sampling rate (44100) and human hearing range (0 - 20000 Hz) give the maximum upper value that a frequency component can take. Sharpness depends on the fundamental frequency in the following way. For harmonic spectra, where each partial is an integer multiple of the fundamental frequency, the maximum number of harmonic partials is determined by dividing the upper limit by the fundamental frequency. For example, a fundamental frequency at 1000 Hz allows a maximum sharpness level of 20000/1000 = 20 harmonic partials to be specified whereas a fundamental frequency at 100 Hz allows 20000/100 = 200 harmonic partials to be specified. Therefore, the range of sounds that can be produced at lower fundamental frequencies is much larger than that of higher fundamentals. It should be noted that in the initial Sound Mosaics prototype, the maximum number of partials that comprise a sound object cannot exceed one hundred. As a result of the above, our texture contrast algorithm has been designed to adapt to the different ranges of sounds that can be produced at different fundamental frequencies. In other words, the number of possible contrast values varies with the number of possible sharpness values.

Sensory dissonance is related to the frequency differences between the sound's frequency components and when these differences are very small then a distinct beating occurs that gives rise to a sensation of sensory dissonance (Sethares 1999). In a series of experiments with pairs of pure tones, Plomp and Levelt (1965) found that sensory dissonance reaches its maximal point at approximately 1/4 of the relative critical bandwidth (see Figure 6.8).

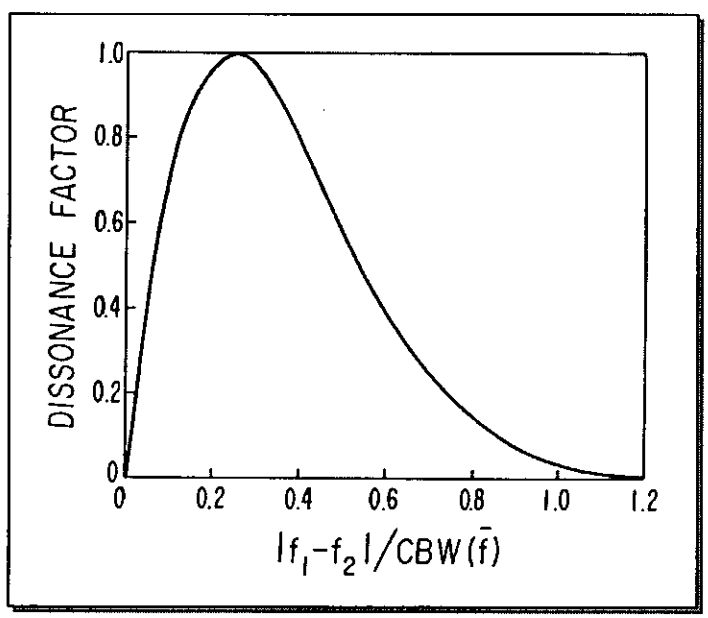


Figure 6.8: Sensory dissonance of two simultaneously sounding pure tones as a function of critical bandwidth (after Plomp and Levelt (1965)).

Hutchinson and Knopoff (1978) extended Plomp and Levelt's research and suggested the following formula for the measurement of sensory dissonance (D) for the case of complex sounds as a sum of the dissonances between all pairs (i,j) of frequency components in the sound's spectrum:

$$D = \frac{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} A_i A_j g_{ij}}{\sum_{i=1}^{N} A_i^2}$$

$$g_{ij} = \left|f_i - f_j\right| / \mathrm{CBW}(\bar{f}) \quad , \quad \mathrm{CBW}(\bar{f}) = 1.72(\bar{f})^{0.65}$$

$A_{i,j}$: Amplitudes of the i_{th} and j_{th} partials
$f_{i,j}$: Frequencies of the i_{th} and j_{th} partials
$\bar{f}$: Mean Frequency of the i_{th} and j_{th} partials

We based our sensory dissonance algorithm on the above formula, although in the current implementation all frequency components have equal amplitudes ($A_{i,j}$) since variation in the amplitude envelope was not included in our model of timbre perception. Therefore, the major factor to control sensory dissonance is the g factor plotted on the horizontal axis in Figure 6.8. Our algorithm is designed to produce eight g values in the range 0.3 - 1.0 (maximum and minimum sensory dissonance respectively) by shifting the harmonic spectrum produced by the sharpness algorithm. However, as the frequency differences between the partials become smaller for larger values of sensory dissonance, an interaction occurs with the perceived level of sharpness, since the position of the spectral centroid as computed by the sharpness formula depends heavily on the frequencies of the partials. In order to address this interaction and allow the independent control of these two dimensions, Sound Mosaics uses the upper limiting frequency of the harmonic spectrum as an indication of the spectrum width and automatically fills in the missing high-frequency content of the sound in order to keep the spectral centroid approximately constant for different levels of sensory dissonance.

Finally, compactness is a dimension of timbre related to the differences between tone-like and noise-like sounds. However, as mentioned in §3.1.3 the formulation of a measurement scale for compactness has been proven difficult (Bismarck 1974a). In Sound Mosaics, once the spectrum is computed for the desired sharpness and sensory dissonance levels, the compactness algorithm produces noise bands centred at the partial frequencies (a similar approach can be found in Barrass (1997)). The noise bandwidths are varying according to the degree of texture CG in order to achieve a tone-noise morphing effect. The initial Sound Mosaics prototype provides eight levels of

compactness with the following noise bandwidths: 10 Hz, 30 Hz, 60 Hz, 90 Hz, 120 Hz, 150 Hz, 200 Hz, and 250 Hz.

6.3 Summary

In this chapter, we presented the initial design and implementation of Sound Mosaics, a prototype user interface for sound synthesis based on direct manipulation of visual representations. Sound Mosaics incorporates a novel method of sound visualisation based on the findings of our empirical investigations as these were discussed in the previous two chapters. Various implementation details have been discussed with regards to the two main components of Sound Mosaics, namely the *image* and *sound* synthesis components.

Overall, our design methodology is based on an iterative design process involving the design and evaluation of prototypes and at this research stage the initial Sound Mosaics prototype serves as a medium to examine the strengths and weaknesses of our approach as discussed in the next chapter.

7

Evaluation I

In this chapter, we present the evaluation framework for our prototype user interface in order to measure the extent to which our research goals and objectives have been attained. The evaluation of visual representations for sound synthesis or related purposes (e.g. sound analysis) is an issue that has generally been disregarded in computer music research. This situation makes it rather difficult to find well-established methods to evaluate a particular visual representation or to compare different representations against the same evaluation criteria. Our research attempts to set the necessary groundwork for future investigations in this area. Furthermore, it is argued that there is a need to view computer-based sound synthesis tools as interactive systems and place them within a design context that can benefit from attempts to design effective interactive systems in other research fields (e.g. Human-Computer Interaction). In particular, measuring the *usability* of a computer-based system can provide important insight on various aspects of the system and inform future design decisions (Newman and Lamming 1995, Noyes and Baber 1999, Paterno' 2000). For these reasons, the evaluation framework for the initial implementation of Sound Mosaics consisted of two controlled experiments in order to answer the following questions:

- How does the visualisation framework employed in Sound Mosaics compare with existing ways of sound visualisation?

- How usable and useful is Sound Mosaics in terms of the chosen design strategy?

7.1 Challenging the Frequency-Domain Paradigm - Part I

The purpose of the first evaluation experiment was to compare our visualisation framework with other frameworks currently used in the context of computer-based sound synthesis. As discussed in more detail in §2.1.1, current visual representations of sound fall into two main classes: *time-domain* and *frequency-domain* representations. Evaluation studies of either time-domain or frequency-domain representations do not appear to have been previously performed in the related literature. Our research is a first attempt to define a set of appropriate criteria in order to investigate the strengths and limitations of current visual representations of sound. In particular, we chose to compare our visualisation framework with the frequency-domain framework since both frameworks are based on spectral information and can represent all the auditory dimensions under examination. Time-domain representations were excluded from this study because of the limitations discussed in §2.1.1 and their inherent inadequacy to represent some auditory dimensions. For example, we cannot tell the pitch of a particular sound by looking into a single period of its waveform.

7.1.1 Method

Experimental Design

We chose a within-subjects design, where each participant performed a series of image-sound association tasks for *both* visualisation frameworks under investigation. The visualisation frameworks were presented in a random order for each subject in order to control possible ordering effects.

Subjects

The subjects were five students in their second year of a bachelor degree in sonic arts and five research students in computer science. All subjects were given the same screening questionnaire used in our previous experiments, extended to include information about any experience with graphic sound synthesis tools (see §A.2 for a sample copy of the questionnaire). The ten participants in this experiment were split into music and non-music groups, each of five subjects, depending on their level of musical experience. None of the subjects had taken part in any of our previous empirical investigations.

Apparatus and Stimuli

The computer application employed in this experiment comprised an image palette and twenty sound sequences (ten sequences for each of the two visualisation frameworks). The content of the image palette varied according to the visualisation framework presented for each task as described later in this section. The design was based on an application used in one of our previous empirical investigations (see Figure 5.3).

Since all tasks had to be performed for both visualisation frameworks, we decided to use two variation modes (either ascending or descending and non-monotonic) for each auditory dimension in order to keep the experiment within reasonable time limits. As a result, two series of sound sequences were designed for each of the five perceptual dimensions of sound thus yielding ten sequences in total. The sound stimuli comprising the sequences were designed with Sound Mosaics and each sequence consisted of five sounds.

The sound stimuli for pitch were pure tones varying in their fundamental frequency at constant amplitude levels. In the case of loudness, we used pure tones with the same fundamental frequency at varying levels of amplitude. For dimensions of timbre, the stimuli were complex tones varying in one timbral dimension while the remaining two were kept constant. All complex tones had the same fundamental frequency and overall amplitude levels in order to approximately equalise pitch and loudness.

As mentioned above, the contents of the image palette varied analogously to the visualisation framework presented for each task. The image stimuli used for Sound Mosaics (see Figure 7.1 and Colour Plate E.5) were visual representations of the sound stimuli used in the experiment and belonged to two classes:

- Colour images that corresponded with sounds changing either in pitch or loudness.

- Texture images that corresponded with sounds changing in any of the three dimensions of timbre.

In order to obtain the visual stimuli for the frequency-domain framework (see Figure 7.2 and Colour Plate E.6) we used *MetaSynth* (U & I Software 1998). As described in §2.1.1, frequency-domain representations created with MetaSynth are drawn on a 2-D plane with the vertical and horizontal axes representing frequency and time respectively. In addition, a black-white scale is used to represent the amplitude (soft-loud) of the individual frequency components. Therefore for pure tones, *height* and *brightness* are the visual dimensions associated with pitch and loudness respectively (i.e. the first two image sequences in Figure 7.2). In the case of complex tones, frequency-domain representations of sound were composed of a series of horizontal line components, where *line addition*, *pixelation* and *density* correspond to auditory sharpness, compactness and sensory dissonance respectively (i.e. the last three image sequences in Figure 7.2). Note that these visual terms are based on our own interpretation of the frequency-domain framework. For example, since sharpness increases with the addition of higher frequency partials line addition seems an appropriate term to use for this dimension of timbre. Similarly, sensory dissonance increases when frequency components are closer to each other thus it is the density of the line components that determines the degree of sensory dissonance. The frequency-domain stimuli were created using the following two methods:

- For sounds varying either in pitch or loudness, the visual stimuli were drawn directly on MetaSynth's *Image Synth* window for the same fundamental frequencies and amplitude levels used in Sound Mosaics.

- For sounds changing in any of the three dimensions of timbre we used MetaSynth's *Filter* function to analyse the corresponding sound stimuli from Sound Mosaics. The Filter function performs a Fast Fourier Transform analysis of sound and produces a frequency-domain visual representation of sound as described above. It should be further noted that the Filter function draws all spectral information in an orange hue, although this is not associated with any particular auditory dimension.

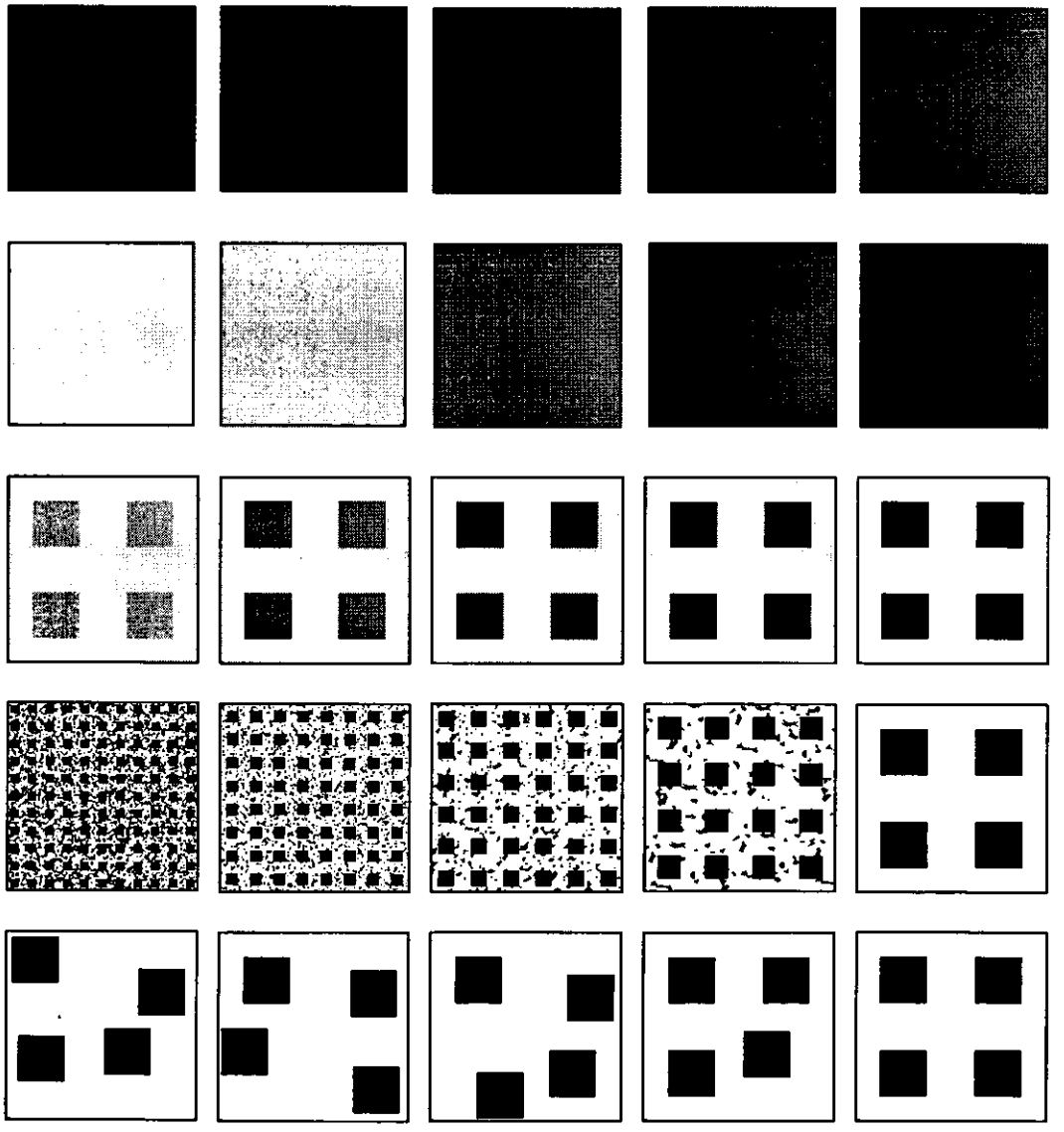


Figure 7.1: The above visual stimuli were used in the evaluation of the initial implementation of Sound Mosaics to form the content of the image palette when subjects were presented with the Sound Mosaics visualisation framework. From top to bottom: Brightness, Saturation, Contrast, Coarseness/Granularity, and Repetitiveness. See also, Colour Plate E.5.

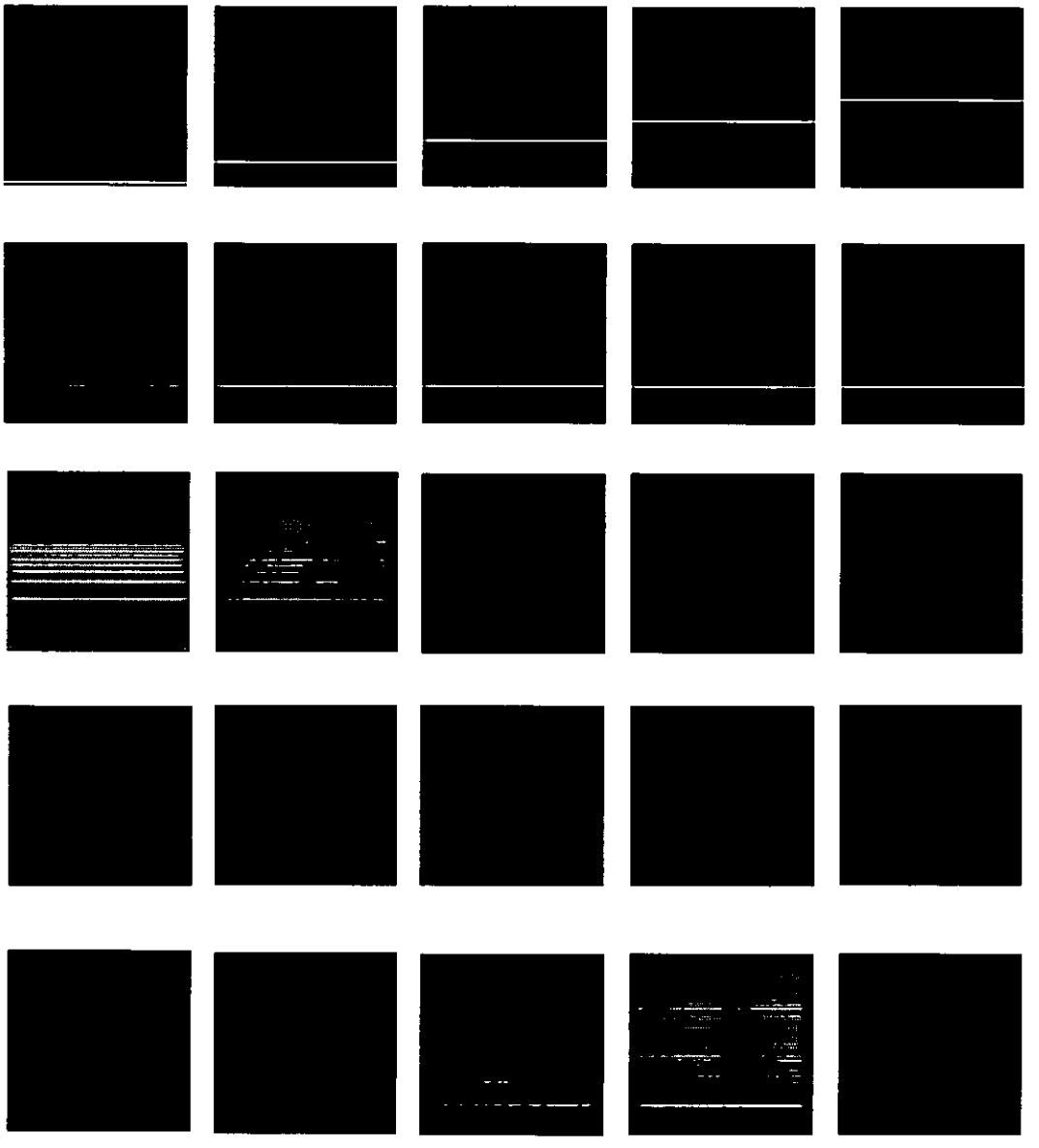


Figure 7.2: The above visual stimuli were used in the evaluation of the initial implementation of Sound Mosaics to form the content of the image palette when subjects were presented with the frequency-domain visualisation framework. From top to bottom: Height, Brightness, Line addition, Pixelation, and Density. See also, Colour Plate E.6.

All frequency-domain images had an original size of 128×256 pixels that was horizontally downscaled to 128×128 pixels in order to match the size of the Sound Mosaics images and fit within one screen when presented to the user. Since there was no variation of the individual frequency components over time (represented on the horizontal axis), horizontal scaling did not affect the appearance of the original visual representations (see Figure 7.3).

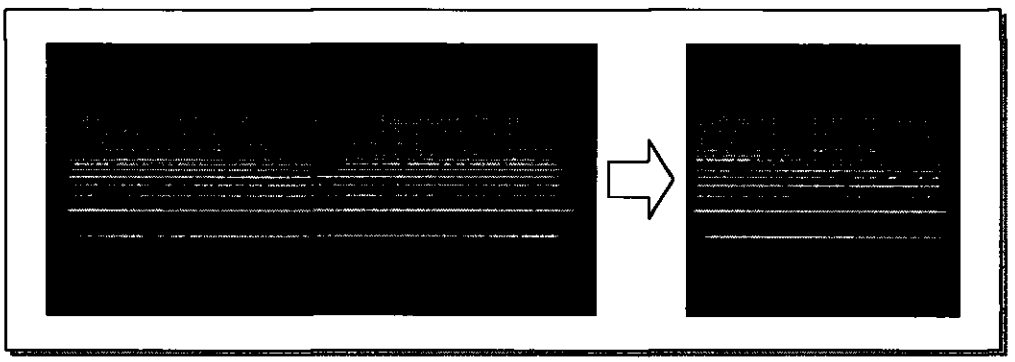


Figure 7.3: An original 128×256 frequency-domain representation (left part) created with MetaSynth downscaled horizontally to a 128×128 image size (right part).

Experimental Task

At the beginning of each session, the experimenter demonstrated how to use the computer-based application and there was a short practice period for subjects to familiarise themselves with the task and the image and sound stimuli incorporated in this experiment. The experimental task was: for the current sequence of five sounds to create a sequence of five corresponding images selected from the current image palette. Before proceeding to the next sound sequence, subjects were instructed to specify their level of confidence for their current sound-image associations by filling a short questionnaire (see §A.3 for a copy of the questionnaire). During the experiment, both image and sound stimuli were introduced in a random order for each subject. Each subject completed the task for twenty sound sequences. The experimenter was present throughout the experiment recording observations that formed the basis for post experiment interviews with subjects. Finally, the application logged subjects' selections in the form of screenshots, as well as completion times per task.

Evaluation Criteria

The main criteria used in this study to evaluate the Sound Mosaics and frequency-domain visualisation frameworks were *comprehensibility* and *intuitiveness*. Comprehensibility was related to the accuracy of the auditory-visual associations

performed by subjects. An auditory-visual association was considered accurate if the subject had selected at least three images from the correct visual dimension as a response to the varying auditory dimension in terms of the corresponding visualisation framework. Table 7.1 shows the auditory-visual associations for both visualisation frameworks.

Dimension	Sound Mosaics	Frequency-Domain
Loudness	Saturation	Brightness
Pitch	Brightness	Height
Sharpness	Contrast	Line Addition
Compactness	CG	Pixelation
S. Dissonance	Repetitiveness	Density

Table 7.1: The target auditory-visual associations for the Sound Mosaics and frequency-domain visualisation frameworks. The visual texture dimension of coarseness-granularity has been abbreviated as CG.

Intuitiveness was related to performance issues such as task completion time and level of confidence. Subjects used a five-point scale to express their level of confidence for each image-sound association task.

7.1.2 Analysis of Results

We now present the results obtained from the above-described experiment for each of our evaluation criteria. For simplicity of presentation Sound Mosaics, frequency-domain, and the visual texture dimension of coarseness-granularity are abbreviated as SM, FD and CG respectively.

Comprehensibility

Tables 7.2 and 7.3 show the overall results obtained by music subjects for sequences that varied in either pitch or loudness for SM and FD representations respectively. In the case of SM, accuracy levels reached 80% and 70% for pitch and loudness respectively. For FD representations, accuracy levels reached 100% and 60% for pitch and loudness respectively. The corresponding results obtained by non-music subjects are presented in Tables 7.4 and 7.5. In the case of SM, accuracy levels reached 90% and 60% for pitch and loudness respectively. For FD representations, accuracy levels reached 80% and 70% for pitch and loudness respectively.

Sound Mosaics	Music Subjects	
Selection Strategy	**Pitch**	**Loudness**
Brightness	80	30
Saturation	20	70
None	0	0
Total (%)	100	100

Table 7.2: Overall results for the Sound Mosaics framework obtained by music subjects for sequences varying in either pitch or loudness.

Frequency-Domain	Music Subjects	
Selection Strategy	**Pitch**	**Loudness**
Brightness	0	60
Height	100	40
None	0	0
Total (%)	100	100

Table 7.3: Overall results for the frequency-domain framework obtained by music subjects for sequences varying in either pitch or loudness.

Sound Mosaics	Non-Music Subjects	
Selection Strategy	**Pitch**	**Loudness**
Brightness	90	40
Saturation	10	60
None	0	0
Total (%)	100	100

Table 7.4: Overall results for the Sound Mosaics framework obtained by non-music subjects for sequences varying in either pitch or loudness.

Frequency-Domain	Non-Music Subjects	
Selection Strategy	**Pitch**	**Loudness**
Brightness	20	70
Height	80	30
None	0	0
Total (%)	100	100

Table 7.5: Overall results for the frequency-domain framework obtained by non-music subjects for sequences varying in either pitch or loudness.

In more detail, the accuracy levels for pitch strongly suggest that the auditory-visual associations used in both frameworks to represent this auditory dimension are very comprehensible. The use of height in FD representations to represent pitch appears to be very strong for both subject groups, a result that was expected since the *low-high* metaphor for pitch representation in music (e.g. the musical stave) is very strong in cultural terms. Nevertheless, the results are very encouraging for the association between pitch and brightness as employed in SM which performed equally well in terms of accuracy. In addition, the association between pitch and brightness is based more on perceptual reality than on cultural factors and as such has cross-cultural validity (Ware 2000). The situation is very similar for the associations between loudness and the visual dimensions of brightness and saturation as used in FD and SM representations respectively, although in both cases accuracy levels were not as high as for pitch. However, a very important paradox can be observed for the sequences varying in loudness. Although subjects associated brightness with loudness when presented with FD representations, they associated loudness with saturation instead of brightness when presented with representations from SM. As discussed in previous chapters of this thesis, the association between loudness and brightness in FD representations is primarily based on the physical correspondence between the two dimensions. Furthermore, brightness as a sensory channel can represent data in an ordered manner and thus it is suitable for the visualisation of ordinal data such as loudness values. However, in the SM framework both brightness and saturation are used for the visualisation of ordered information. Therefore it can be argued that subjects' preference of a saturation-loudness association can be explained either as a preference of saturation over brightness or as a direct association between saturation (strength) and loudness. We suspect that it is this latter similarity between loudness and saturation that subjects used in their judgements. This provides further evidence that although various analogies can be drawn based on physical correspondences between auditory and visual dimensions, perceptual reality might suggest otherwise.

In Tables 7.6 and 7.7 we present the overall results obtained by music subjects for sequences that varied in any of the timbral dimensions. The results for SM representations show a very low accuracy level for the dimensions of sharpness and sensory dissonance (20% and 10% respectively), although the accuracy level for the dimension of compactness reached 100%. For FD representations, accuracy of association reached 10% for sharpness, 50% for compactness, and 70% for sensory dissonance. Furthermore, Tables 6.8 and 6.9 show the overall results obtained by non-music subjects for sequences that varied in any of the timbral dimensions. Accuracy levels for SM representations reached 40%, 90%, and 40% for sharpness, compactness and sensory dissonance respectively. For FD representations, accuracy levels reached 50% for sharpness and 80% for both compactness and sensory dissonance.

Sound Mosaics	Music Subjects		
Selection Strategy	**Sharpness**	**Comp/ness**	**S. Dissonance**
Contrast	20	0	0
CG	50	100	90
Repetitiveness	0	0	10
Mixed	10	0	0
None	20	0	0
Total (%)	100	100	100

Table 7.6: Overall results for the Sound Mosaics framework obtained by music subjects for sequences varying in any of the dimensions of timbre.

Frequency-Domain	Music Subjects		
Selection Strategy	**Sharpness**	**Comp/ness**	**S. Dissonance**
Line Addition	10	0	10
Pixelation	0	50	0
Density	50	40	70
Mixed	20	0	0
None	20	10	20
Total (%)	100	100	100

Table 7.7: Overall results for the frequency-domain framework obtained by music subjects for sequences varying in any of the dimensions of timbre.

Sound Mosaics	Non-Music Subjects		
Selection Strategy	**Sharpness**	**Comp/ness**	**S. Dissonance**
Contrast	40	10	10
CG	60	90	50
Repetitiveness	0	0	40
Mixed	0	0	0
None	0	0	0
Total (%)	100	100	100

Table 7.8: Overall results for the Sound Mosaics framework obtained by non-music subjects for sequences varying in any of the dimensions of timbre.

Frequency-Domain	Non-Music Subjects		
Selection Strategy	**Sharpness**	**Comp/ness**	**S. Dissonance**
Line Addition	50	0	10
Pixelation	0	80	0
Density	40	10	80
Mixed	10	10	10
None	0	0	0
Total (%)	100	100	100

Table 7.9: Overall results for the frequency-domain framework obtained by non-music subjects for sequences varying in any of the dimensions of timbre.

For sound sequences varying in any of the dimensions of timbre, FD representations appear to have performed better than SM. The results were very poor for both frameworks as far as the dimension of sharpness is concerned. The association between compactness and the dimension of CG was very strong in SM for both subjects groups. The association between compactness and pixelation in FD representations was strong for non-music subjects although the results obtained from music subjects were not satisfactory. SM has also performed rather poorly for sound sequences varying in sensory dissonance. In particular, the results suggest that subjects from both subject groups used the dimension of CG to associate variations in any of the dimensions of timbre. Conversely, accuracy levels for FD representations were satisfactory as far as the dimension of sensory dissonance is concerned.

In general, a surprising observation is that non-music subjects gave higher accuracy levels for both visualisation frameworks than subjects with strong musical experience. Another surprising observation is the stark difference between the results obtained in this study for SM and the results of our empirical investigation described in §5.2. These observations led us to examine more closely the two experiments in order to identify the reasons for Sound Mosaics' inadequacy to represent dimensions of timbre using the visualisation framework discussed in §6.1. An important difference between the two empirical studies was that the sound stimuli used in our previous experiment were much simpler in terms of their frequency content. The reader is reminded that for both sharpness and sensory dissonance, sound stimuli had a maximum of six frequency components. However, the sound stimuli used in the evaluation experiment were much more complex involving sounds containing up to 70 partials. As a result, the sound stimuli that comprised the sharpness and sensory dissonance sequences were perceived as complex, a fact that was also confirmed when we examined subjects' responses during the post-experiment interviews. When asked why they used only the composite dimension of CG to represent changes in dimensions other than compactness, the majority of subjects responded that the rest of the SM visual representations were too *simple* to represent the perceived complexity of the sounds. In addition subjects who correctly associated sensory dissonance with texture repetitiveness reported that they found it particularly difficult to perceive the order of the texture images. We believe that this latter issue might have turned away subjects from using images with different levels of texture repetitiveness.

Although accuracy levels for sensory dissonance were higher for FD representations, we strongly suspect that it was not line density that the subjects associated with sensory dissonance but other visual properties of the image stimuli such as the overall image brightness. For example, visual examination of the FD stimuli (see Figure 7.2 and Colour Plate E.6) for the dimensions of sharpness and sensory dissonance indicates that the overall brightness of those stimuli is noticeably different in both cases. As previously mentioned the obtained results for sharpness were rather poor and therefore they need not concern us here. However, in the case of sensory dissonance we performed further analysis of the results based on how close subjects' responses were to the target sound stimuli in terms of the density and brightness of the respective image stimuli. Figures 7.4 and 7.5 show mean responses for music and non-music subjects respectively. It can be noticed that for both subject groups, mean responses for brightness were closer to the target sensory dissonance stimuli than mean responses for density. These results suggest that it was primarily the overall brightness of the images and not the density of the line components that subjects used for the association with sensory dissonance.

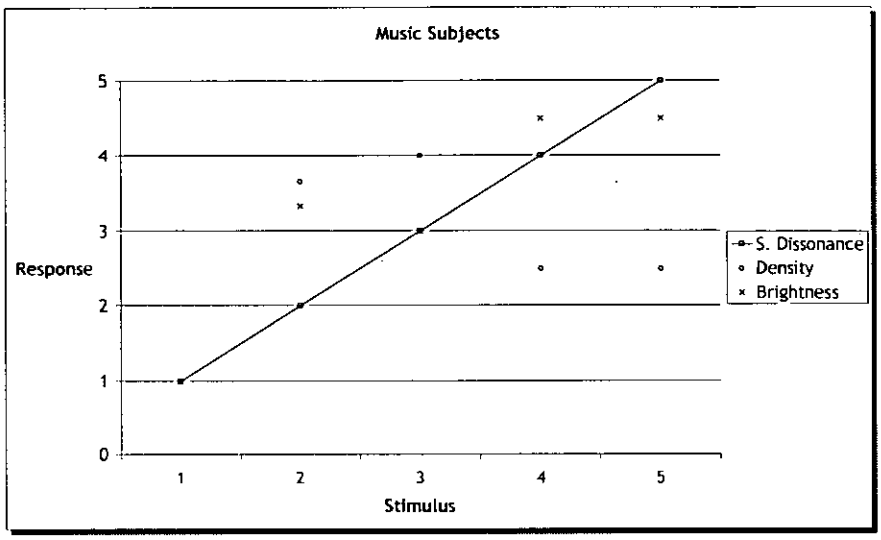


Figure 7.4: Average density and brightness responses obtained by music subjects for sensory dissonance sequences.

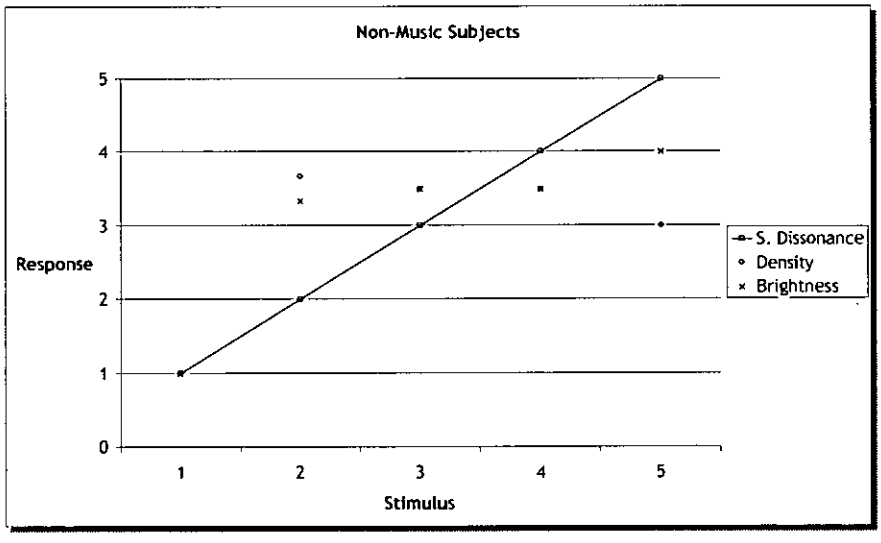


Figure 7.5: Average density and brightness responses obtained by non-music subjects for sensory dissonance sequences.

Intuitiveness

Figures 7.6 and 7.7 show completion times per task for both visualisation frameworks as obtained by music and non-music subjects respectively. In addition, confidence levels per task for both subject groups are shown in Figures 7.8 and 7.9.

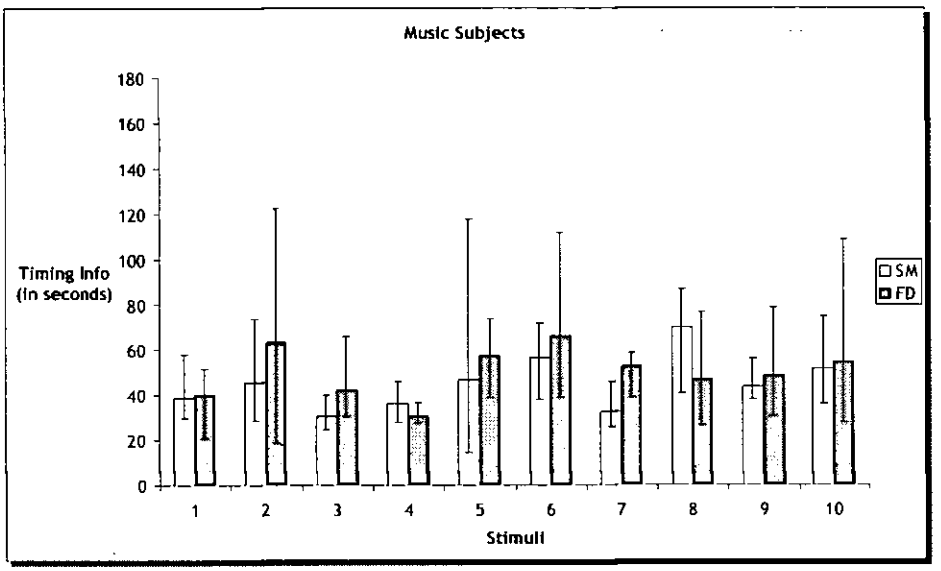


Figure 7.6: Maximum, minimum, and mean completion times per task obtained by music subjects for both visualisation frameworks.

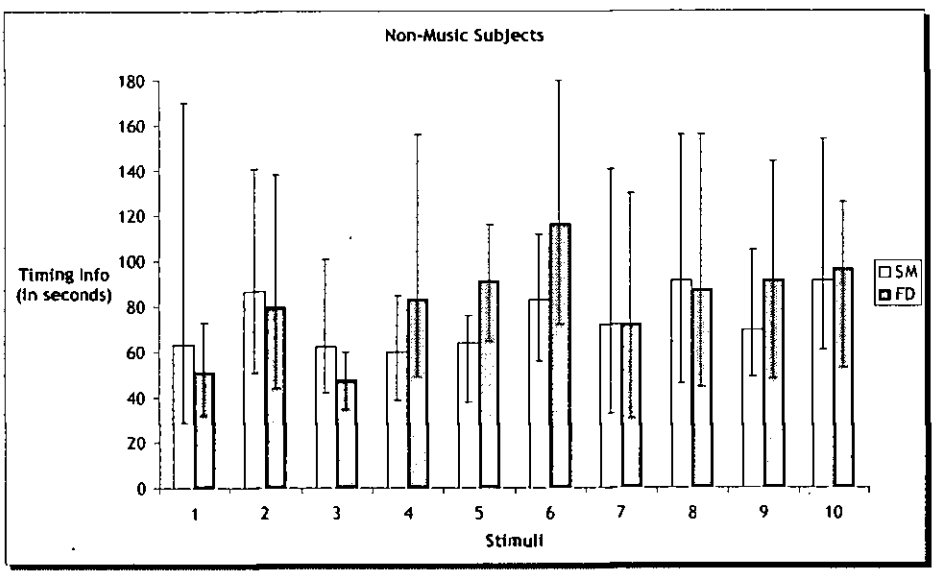


Figure 7.7: Maximum, minimum, and mean completion times per task obtained by non-music subjects for both visualisation frameworks.

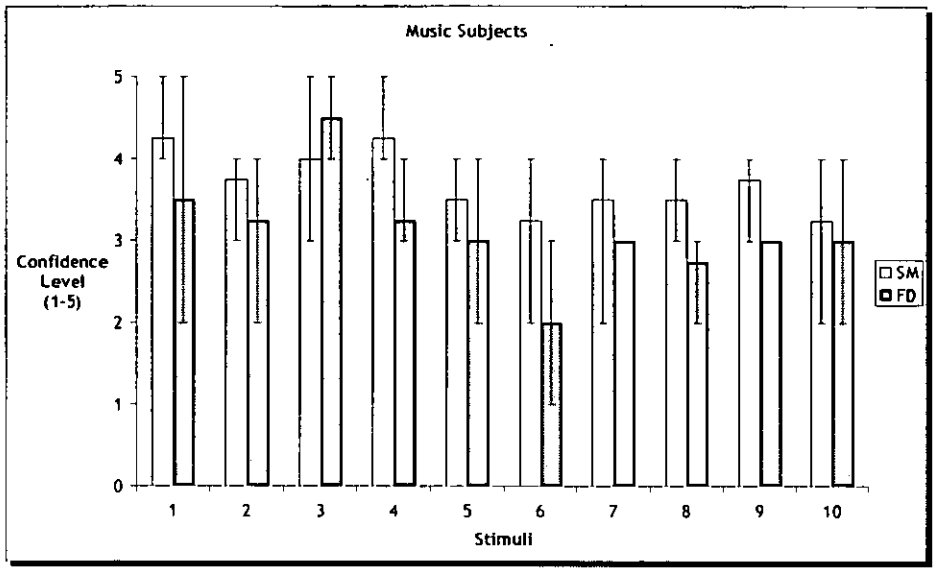


Figure 7.8: Maximum, minimum, and mean confidence levels per task obtained by music subjects for both visualisation frameworks.

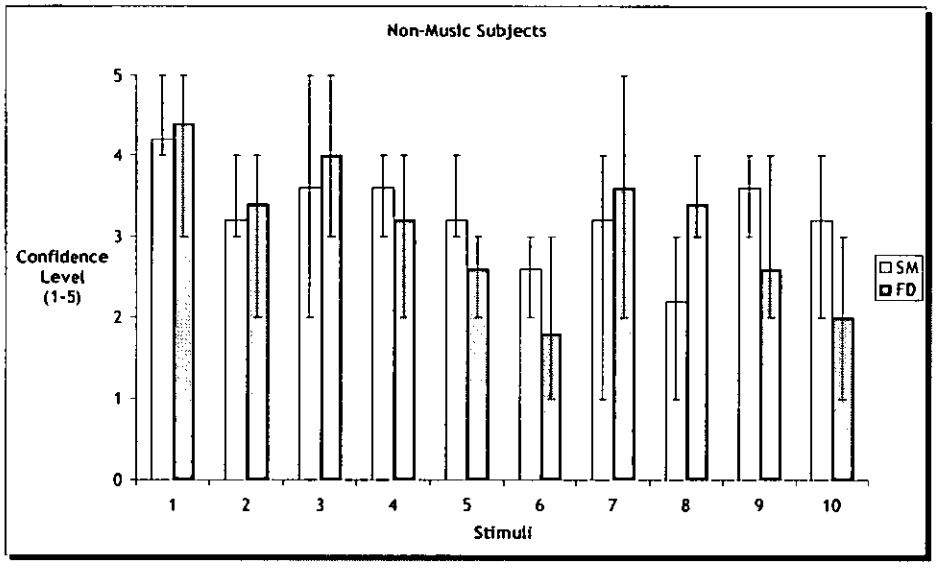


Figure 7.9: Maximum, minimum, and mean confidence levels per task obtained by non-music subjects for both visualisation frameworks.

In general, music subjects were faster than non-music subjects for both visualisation frameworks suggesting that musical experience played an important role in subjects' responses. It can be argued that music subjects were more familiar with the auditory dimensions incorporated in the experiment and thus needed less time to perform the tasks. Music subjects were noticeably faster for SM representations in 5/10 cases (stimuli 2, 3, 5, 6, 7) compared to only one case where FD was faster (stimulus 8) and 4/10 cases

where the two frameworks appear to be equal (stimuli 1, 4, 9, 10). Similarly, non-music subjects were faster for SM representations in 4/10 cases (stimuli 4, 5, 6, 9) compared to 2/10 cases where FD was faster (stimuli 1, 3) and 4/10 cases where the two frameworks appear to be equal (stimuli 2, 7, 8, 10).

Although accuracy levels were in general higher for frequency-domain representations than for Sound Mosaics, it appears that subjects felt more confident creating sequences with SM than FD visual representations (this is more evident for music subjects). Finally, at the end of each session, we further asked subjects to state their preference for either of the two visualisation frameworks. Four out of 5 music subjects expressed a preference for FD in the case of sequences varying either in pitch or loudness. However, in the case of timbral dimensions, 3/5 music subjects preferred SM as opposed to 2/5 subjects who preferred the FD framework. Three out of 5 non-music subjects showed a preference for SM in the case of all auditory dimensions as opposed to only one subject who preferred FD.

The above results indicate that in terms of perceived intuitiveness, Sound Mosaics performed better than frequency-domain representations.

7.2 Usability Evaluation of Sound Mosaics - Part I

An important stage in the development of interactive systems is the evaluation of various aspects of the user interface in order to measure the system's *usability* and identify design choices that need modification and revision. This is particularly important for systems like Sound Mosaics that are at an experimental stage and evaluation is an essential part of the iterative design process (formative evaluation). The evaluation of an interactive system is usually carried out by building and testing a prototype with real users in a controlled environment measuring various usability aspects. The International Standards Organisation (ISO) defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO 9241-11 1998). Based on this definition, usability refers to the aspects of *effectiveness*, *efficiency*, and *satisfaction*. Effectiveness is the accuracy and completeness with which users achieve specific goals. The resources spent on achieving these goals give an indication of the system's efficiency (e.g. learning time, task completion time). Satisfaction is a subjective measure that indicates the users' comfort with and positive attitudes towards the use of the system. Subjective satisfaction is usually measured by a short questionnaire that is given to users at the end of an evaluation session. In the case of Sound Mosaics, we have performed a formative evaluation study to test the design choices incorporated in the initial implementation

described in §6.2 with intended users performing a series of sound synthesis tasks as described in the remainder of this section.

7.2.1 Method

Experimental Design

The intended users of Sound Mosaics are expected to have a substantial interest in music compositional processes. It is anticipated that such users have a musical background either in computer music (e.g. using music programming languages, other sound synthesis applications) and/or in traditional music (e.g. playing musical instruments, reading music, understanding music theory). However, in this evaluation study we were interested to see the influence of musical experience on subjects' performance with Sound Mosaics. To this end, we chose a between-subjects design, where the same Sound Mosaics prototype was used by two groups of subjects (music and non-music) to perform a series of experimental tasks.

Subjects

We had ten participants in this experiment assigned into music and non-music groups, each of five subjects, based on their level of musical experience (screened with the questionnaire in §A.2). Two subjects had previously taken part in the empirical investigation described in §5.2 whereas one subject had taken part in our first empirical study described in §4.2. We didn't expect any influence on the performance of those subjects since the evaluation study was designed to measure the system's usability and not the validity of the underlying visualisation framework that was addressed in our previous empirical studies.

Experimental Task and Stimuli

At the beginning of each session, the experimenter demonstrated how to use the Sound Mosaics prototype and there was a short practice period for subjects to familiarise themselves with the user interface and the experimental task. During the demonstration, the underlying auditory-visual associations were briefly explained to subjects. The Sound Mosaics prototype used in the experiment is shown in Figure 7.10. It can be seen that the interface was slightly modified in order to accommodate the experimental task. For each task, subjects first had to compare two sounds that differed in only one auditory dimension (not known to subjects) and then adjust the settings of the first (or default) sound until they believed there was a good match with the second (or target) sound. The comparison was possible by clicking the "Compare" button, which played the two sounds in succession (default-target) with a 0.5 seconds silence in between and subjects were

allowed to compare the two sounds as many times as they wished. Subjects had to click the "Create" button each time they have changed the settings in order for Sound Mosaics to process the changes and synthesise the sound that would then take the place of the first sound in the comparison. Subjects could submit their final choices via the "Submit"

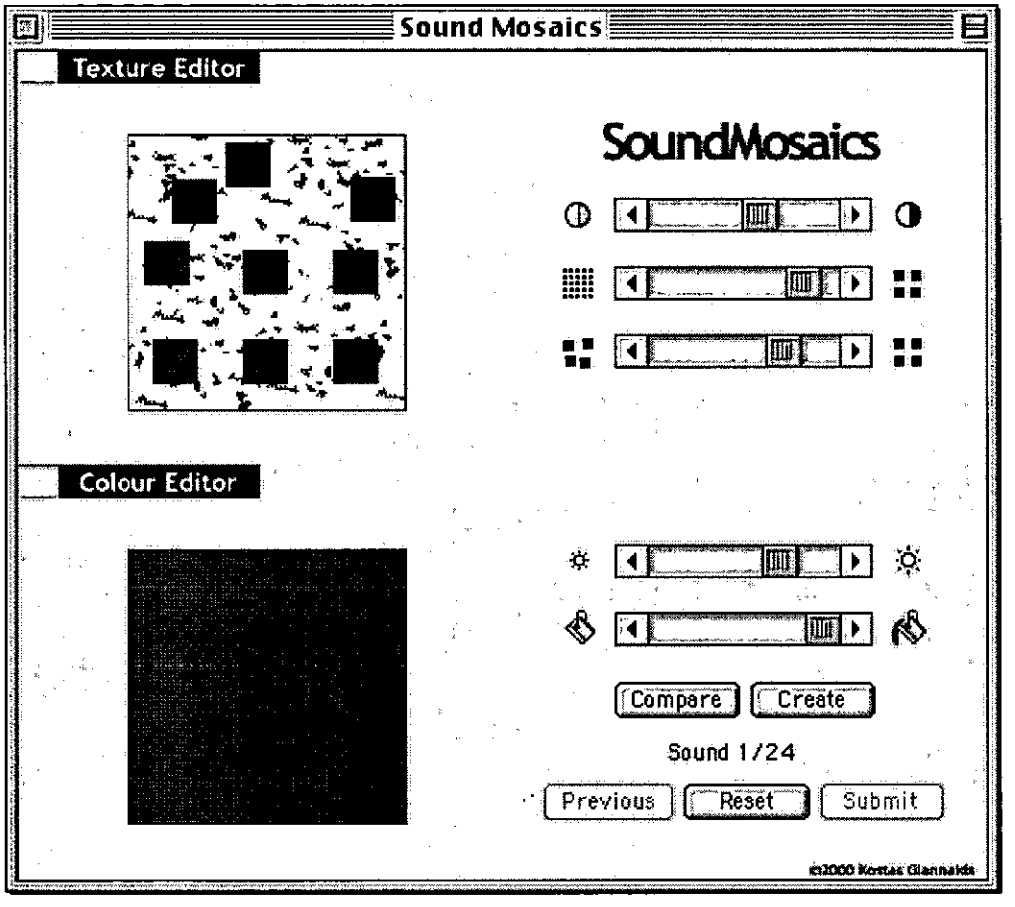


Figure 7.10: The modified Sound Mosaics prototype used in this evaluation study.

button and proceed to the next task. The "Previous" button was solely for the experimenter's use in case unexpected errors (e.g. users accidentally clicking the "Submit" button) required a particular task to be repeated without having to restart the application. In order to reduce the complexity of the task, subjects could click the "Reset" button to reset the settings of the first sound if they felt it deviated significantly from the target sound.

The above-described task was repeated for twenty-four target sounds, of which the first five were part of the training session and the remaining nineteen were the actual experimental stimuli. The nineteen stimuli were composed in the following manner:

- Two target stimuli for the auditory dimension of pitch.

- Two target stimuli for the auditory dimension of loudness.
- Five target stimuli for each of the three perceptual dimensions of timbre.

An initial hypothesis was that subjects' performance would improve faster for the dimensions of pitch and loudness than for any of the dimensions of timbre. For this reason, the emphasis was put on the dimensions of timbre expecting to see an improvement in subject's performance after five experimental tasks for each dimension as opposed to only two for each of pitch and loudness. The order of the tasks was different for each subject in order to avoid ordering effects. Subjects performed the experiment at their own pace and times ranged from fifty minutes to one hour. However, subjects were told at the beginning of the session that the experimenter could ask them to submit their choices after a certain period of time in order to keep the experiment within the 1-hour limit. The experimenter was present throughout the experiment recording observations that formed the basis for post experiment interviews with subjects. Finally, a data collection program logged completion times for each task, subjects' responses in terms of the five auditory and visual dimensions, as well as the number of created sounds and the number of times that subjects compared the default-target sets.

7.2.2 Analysis of Results

This section presents the results obtained from the above-described experiment for each of the usability factors under investigation.

Effectiveness

The main indicator of effectiveness was the quality of subjects' responses, i.e. the closeness of the response to the target stimulus for each task. Table 7.10 presents the five-point scale used to grade each task based on Frøkjær, Hertzum, and Hornbæk (2000).

Score	Description
Very High (<20%)	Brilliant answer
High (20%-40%)	Good and adequate answer
Medium (40%-60%)	Reasonable but incomplete answer
Low (60%-80%)	Inadequate or partially wrong answer
Very Low (>80%)	Failure, a completely wrong answer

Table 7.10: The five-point scale used in this study to grade the quality of subjects responses based on Frøkjær, Hertzum, and Hornbæk (2000). Percentage values indicate the degree of deviation from the target.

In Tables 7.11 and 7.12, we present the scores obtained from music and non-music subjects respectively, for all 19 sound stimuli classified according to each auditory dimension. The scores are based on the average subjects' responses for each stimulus as shown in Figures 7.11 - 7.15. For each auditory dimension, these figures show both music and non-music subjects' mean responses as well as the target and default levels for each stimulus (figures showing subjects' responses for all dimensions in each of the 19 stimuli can be found in Appendix B).

Score	L	P	S	C	SD	Total
V. High	2	2	3	3	3	13
High	0	0	2	1	2	5
Medium	0	0	0	1	0	1
Low	0	0	0	0	0	0
V. Low	0	0	0	0	0	0
Total	2	2	5	5	5	19

L: Loudness, P: Pitch, S: Sharpness,
C: Compactness, SD: S. Dissonance

Table 7.11: Overall scores for each auditory dimension obtained by music subjects.

Score	L	P	S	C	SD	Total
V. High	1	1	2	4	0	8
High	1	0	2	1	1	6
Medium	0	1	0	0	1	2
Low	0	0	1	0	2	3
V. Low	0	0	0	0	1	1
Total	2	2	5	5	5	19

L: Loudness, P: Pitch, S: Sharpness,
C: Compactness, SD: S. Dissonance

Table 7.12: Overall scores for each auditory dimension obtained by non-music subjects.

As expected, music subjects performed better than non-music subjects. However, in both cases, the sum of *high* and *very high* scores represents the vast majority of the sound stimuli (18/19 or 94.7% for music subjects and 14/19 or 73.7% for non-music subjects). Non-music subjects appear to have had particular difficulty with stimuli varying in sensory dissonance. Our speculation is that the difficulty associated with sensory dissonance was a function of three main factors. The first factor is possible interaction

between pitch and sensory dissonance, which caused subjects to vary both pitch and sensory dissonance leading to poor matching results with the target stimuli. The second factor may be attributed to the lack of good correspondence between the variation in texture repetitiveness and variation in sensory dissonance. A third factor may be non-music subjects' unfamiliarity with the auditory dimension of sensory dissonance. Neither subject group appear to have encountered difficulties with the dimensions of sharpness and compactness.

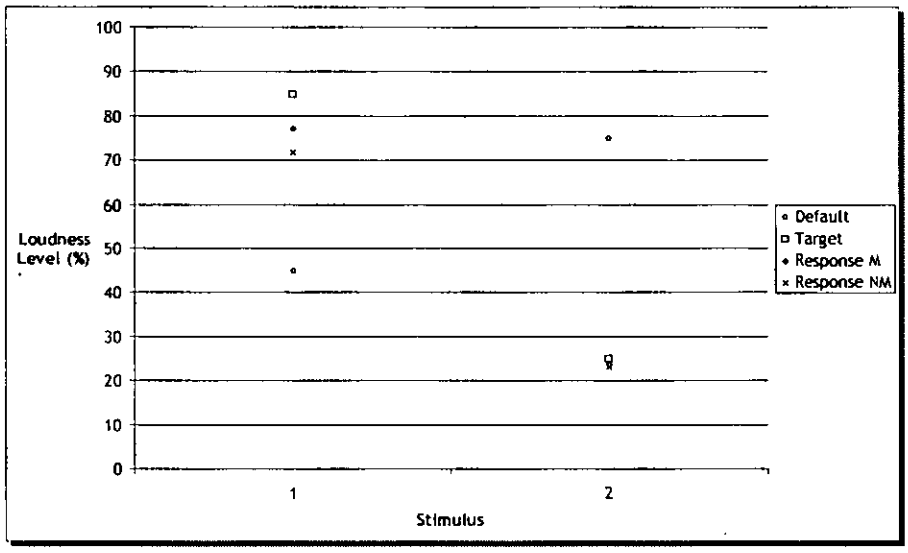


Figure 7.11: Average results obtained by both subject groups for loudness stimuli.

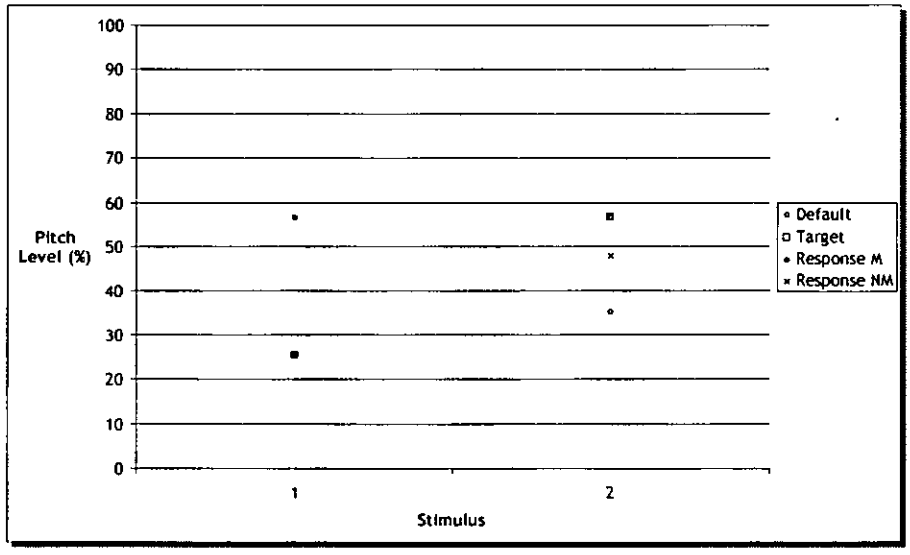


Figure 7.12: Average results obtained by both subject groups for pitch stimuli.

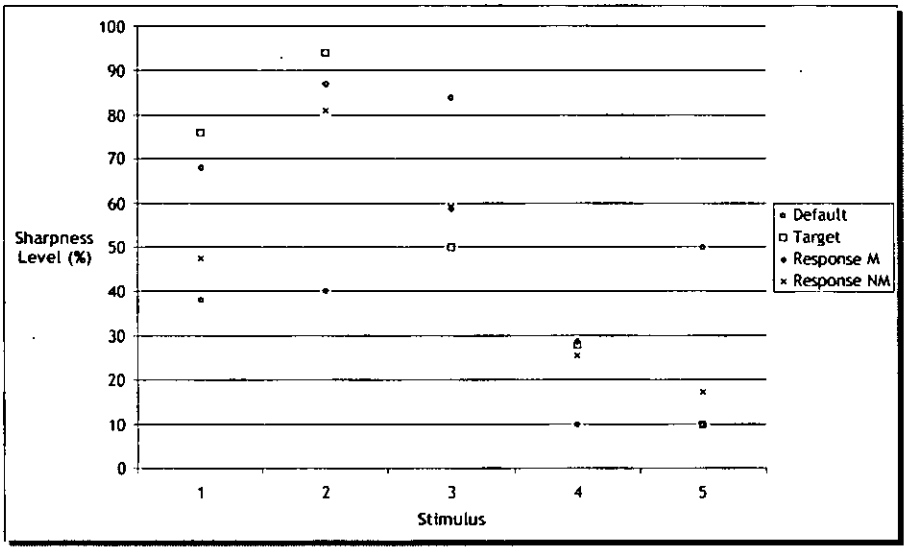


Figure 7.13: Average results obtained by both subject groups for sharpness stimuli.

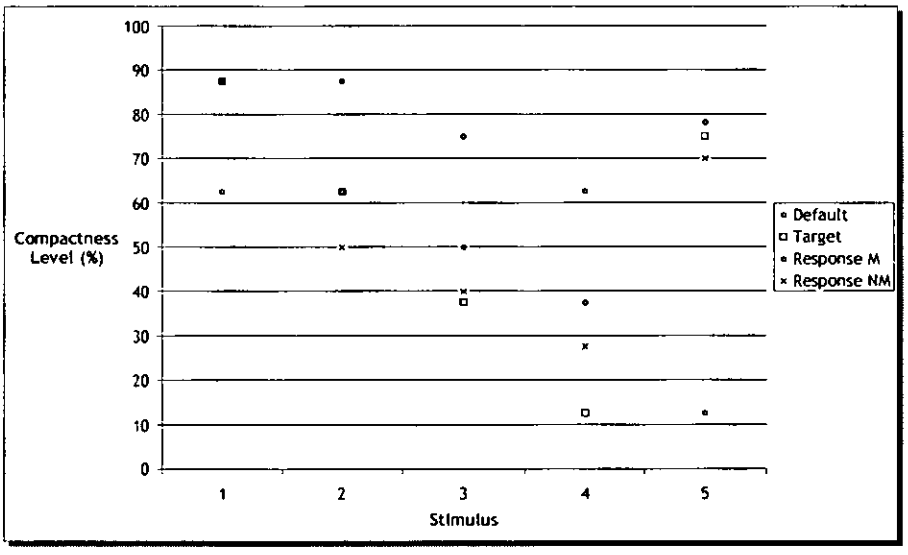


Figure 7.14: Average results obtained by both subject groups for compactness stimuli.

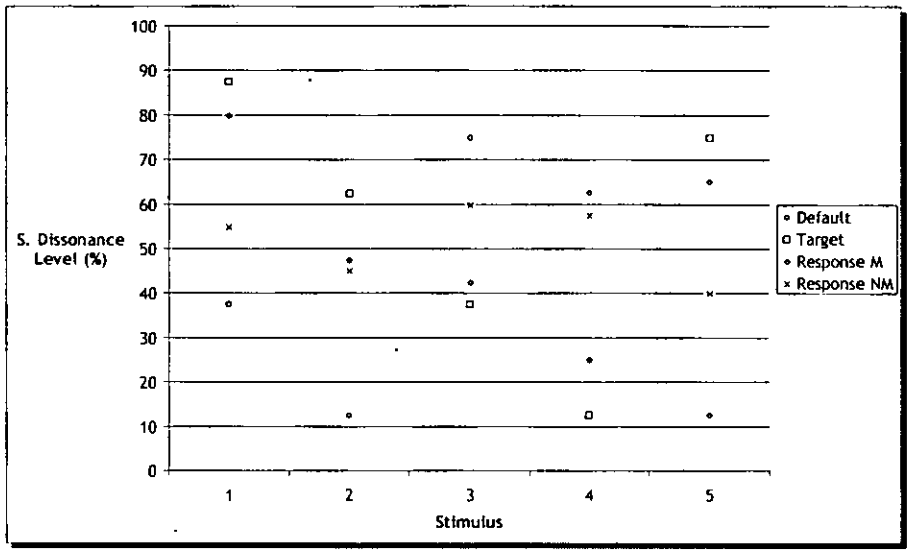


Figure 7.15: Average results obtained by both subject groups for sensory dissonance stimuli.

Efficiency

We used task completion time as the main indicator of efficiency. The longest time that subjects were allowed to spend on a particular stimulus was three minutes. After that period of time, the experimenter asked the subjects to submit their current choices and proceed to the next task. We expected to see a decrease in task completion times, as subjects became more familiar with the sounds on a particular dimension and expected a possible target of one minute or less per stimulus. Figures 7.16 and 7.17 show the results for music subjects. The data displayed are minimum, maximum and average times achieved for each stimulus in each of the five auditory dimensions. Average completion times for pitch and loudness were short (<60 seconds) confirming our expectations although the large variability in loudness stimuli indicates possible problems caused by colour discrimination issues. Decreasing trends can be noticed for all three dimensions of timbre (more evident in the compactness stimuli) with average times around the 1-minute target. The situation was very similar for non-music subjects and the corresponding results are shown in Figures 7.18 and 7.19. The outlying compactness stimulus 5 in Figure 7.19 may be attributed to the complexity of the respective sound stimulus (possible interaction between different dimensions).

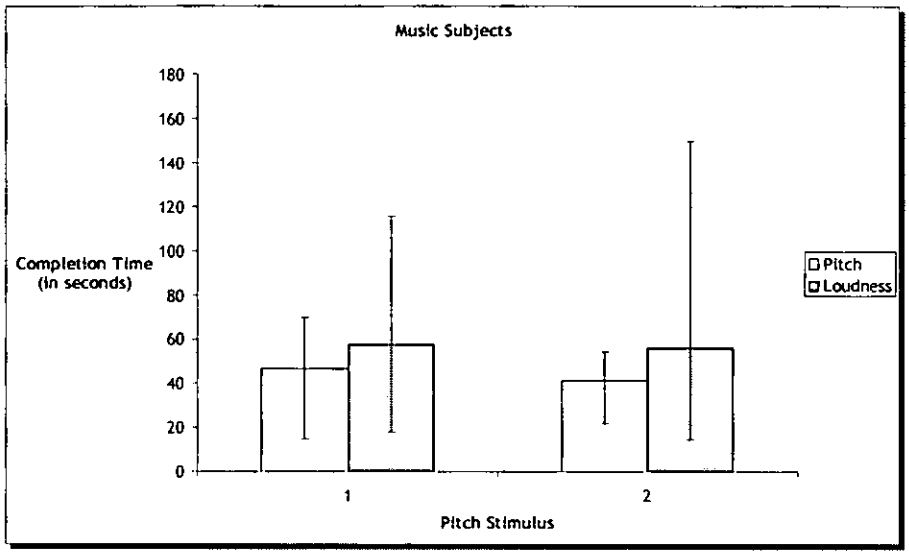


Figure 7.16: Maximum, minimum, and mean completion times obtained by music subjects for pitch and loudness stimuli.

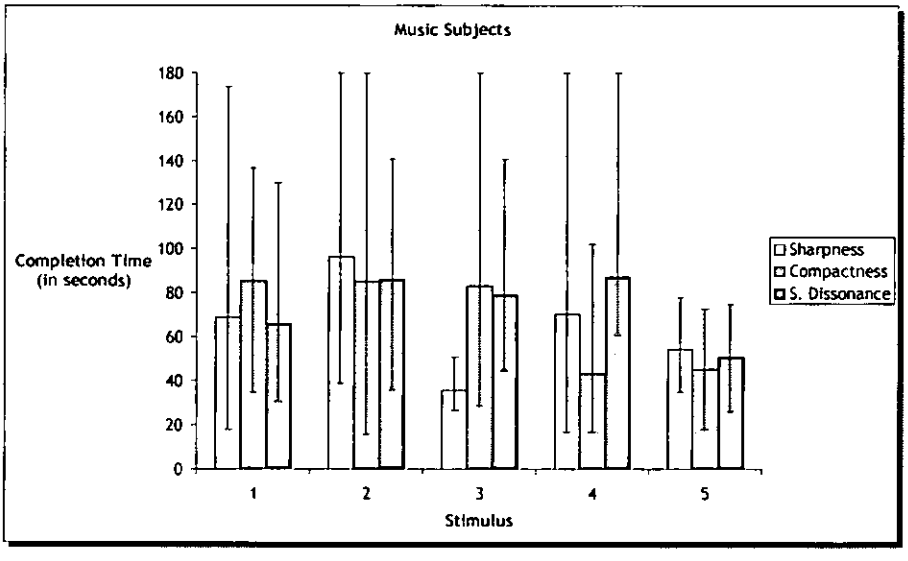


Figure 7.17: Maximum, minimum, and mean completion times obtained by music subjects for timbre stimuli.

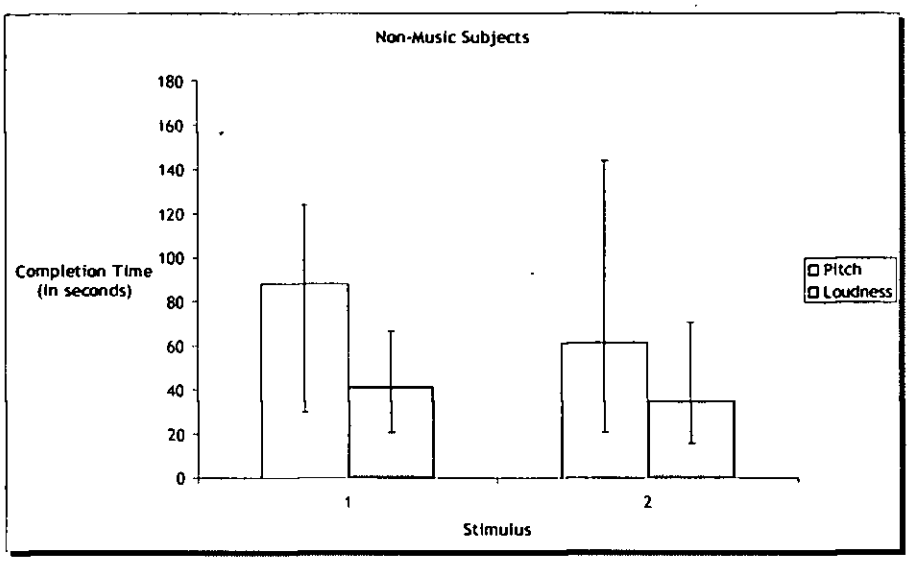


Figure 7.18: Maximum, minimum, and mean completion times obtained by non-music subjects for pitch and loudness stimuli.

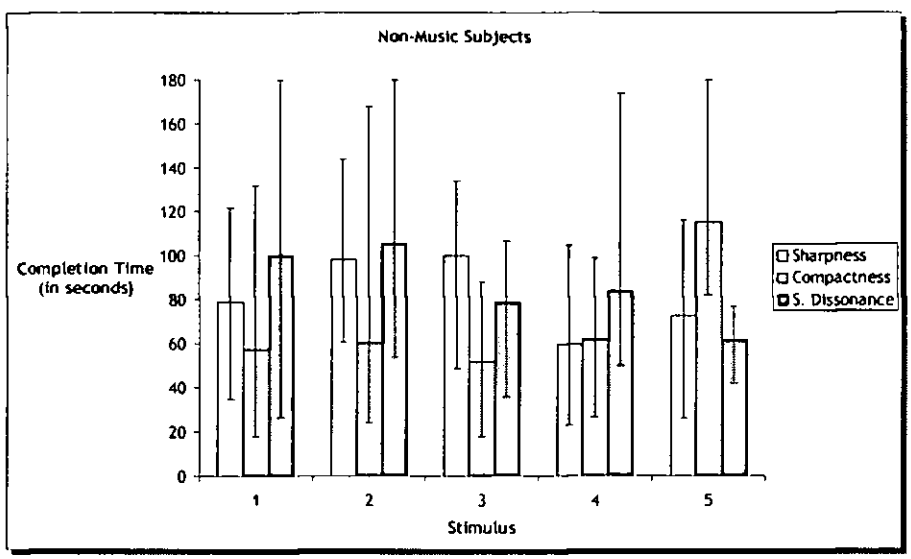


Figure 7.19: Maximum, minimum, and mean completion times obtained by non-music subjects for timbre stimuli.

Subjective Satisfaction

In order to measure subjects' satisfaction with Sound Mosaics we used the Subjective Users Satisfaction (SUS) questionnaire (see §A.4 for a copy of SUS) by Brooke (1996) and the scores for each subject are presented in Table 7.13. The mean scores obtained by music and non-music subjects were 69% and 67% respectively suggesting that subjects were satisfied with the various aspects of Sound Mosaics.

SUS Results	Music	Non-Music
Subject 1	67.5	72.5
Subject 2	62.5	75
Subject 3	67.5	55
Subject 4	75	67.5
Subject 5	72.5	65
Average (%)	69	67

Table 7.13: Subjective satisfaction scores obtained by music and non-music subjects for the initial Sound Mosaics prototype.

Discussion

Based on the above-presented results, it can be concluded that Sound Mosaics achieved very satisfactory levels of usability in terms of *effectiveness*, *efficiency* and *subjective satisfaction*. However, the significance of this conclusion is limited due to an undesirable situation that was observed during the evaluation sessions. In more detail, most subjects reported (quite often after a few minutes into the session) that they concentrated their attention on using the scrollbar objects as means of getting visual feedback for the current values of the auditory parameters incorporated in Sound Mosaics and not the colour and texture images. In other words, users shifted their attention from the colour and texture editors to the scrollbar controls, as these seemed to provide a better representation of the values of auditory dimensions. Therefore, it might be argued that in the initial Sound Mosaics prototype, the colour and texture images were redundant. Our observations during the experiment and the users' responses in post-experiment interviews led us to conclude that this undesirable situation was a function of three factors.

The first factor is related to the design and visual presentation of the scrollbar objects. In more detail, the scroll box was visible indicating the current value for a particular visual dimension. Therefore, it was always possible to guess the value for a particular dimension (e.g. high, low, higher than before, etc.) just by looking at the position of the

scroll boxes since values were meaningfully ordered from left to right in a low-high fashion. The second factor is related to the limited perceptual discrimination for certain auditory and visual dimensions. In particular, the number of just-perceivable differences for brightness and saturation appears to be less than the number of steps provided in the initial implementation of Sound Mosaics. This means that although subjects perceived a difference in auditory terms, there was a non-perceivable change in the colour or texture images. However, there was a perceivable change on the scrollbars' visual states as the scroll boxes changed position according to the users' interaction with the system. Subjects also reported that noticeable differences for loudness did not correspond well to the different levels provided in Sound Mosaics. Finally, a third factor might be attributed to the design strategy of having two visual representations for a sound object instead of a single representation of sound.

7.3 Conclusion

This chapter presented the evaluation framework for our sound visualisation method and the initial implementation of Sound Mosaics described in §6.2. The evaluation consisted of two parts. The first part dealt with a comparison study of the Sound Mosaics and frequency-domain visualisation frameworks along the criteria of comprehensibility and intuitiveness. In terms of comprehensibility, although the results suggested that both frameworks performed satisfactorily and equally well for the auditory dimensions of pitch and loudness, the results for the perceptual dimensions of timbre were poor (with the exception of compactness) in both cases. As far as the criterion of intuitiveness is concerned, response times and confidence levels obtained by the participants in our experiment indicated that Sound Mosaics representations were more intuitive than frequency-domain representations. However, this conclusion when considered in conjunction with the above comprehensibility results is of limited importance.

The second part of our evaluation was a usability study of the initial Sound Mosaics prototype. Although Sound Mosaics scored very well in terms of effectiveness, efficiency, and subjective satisfaction, our observations and post-experiment interviews with subjects led us to conclude that the potential gains from using the visual representations in the colour and texture editors were not adequately investigated.

Overall, our evaluation of Sound Mosaics pointed out the importance of incorporating formative evaluation stages during the design of interactive systems such as computer-based sound synthesis tools. It assisted us in the identification of various problematic issues that need to be resolved in order to achieve better results. In the next chapter, we present our attempt to address the limitations of our initial design choices.

8

Sound Mosaics II & Evaluation II

In this chapter, we present a revised implementation of Sound Mosaics based on the insights gained from the evaluation of the initial implementation described in the previous chapter. In the first section, we give detailed descriptions of how we attempted to overcome the limitations of our initial design choices. In the second section, we present the results of a second evaluation study in order to measure the extent to which the revised visualisation framework and design choices improved the performance of Sound Mosaics.

8.1 Revising our Visualisation Framework and Design Choices

8.1.1 The Shift-of-Focus Problem

The initial implementation of Sound Mosaics represented sound visually in two different ways:

- First, as a set of images displayed in the colour and texture editor panels.
- Second, as a set of interface elements, in this case scrollbar values.

One of the insights gained during the evaluation of our initial implementation of Sound Mosaics was that users concentrated their attention on using the scrollbar objects as means of getting visual feedback for the current values of the auditory parameters incorporated in Sound Mosaics. For the purposes of our research, this *shift-of-focus* was an undesirable feature since we were interested in investigating the extent to which users can absorb information about auditory percepts solely from the colour and texture images in a useful and usable manner. There are various reasons why visual representations such as the colour and texture images are preferable to scrollbar controls and these can be outlined as follows:

- The properties of colour and visual texture are closer to the mapping domain than those of the scrollbars. As a result, the auditory-visual associations are clearer.
- The colour and texture images could also make the relationships between different perceptual dimensions more visible.
- Visual representations that take advantage of sensory channels such as colour and visual texture are fast to grasp and easy to remember.
- They are also more 'economical' in the sense that they provide more information in less space. This is an important issue for future

implementations that will incorporate more dimensions, however increasing the number of scrollbars will hinder the way users interact with the system.

- Finally, these representations could be more engaging for users, thus contributing to higher levels of subjective satisfaction.

The solution that we suggested for the above-described problem was to eliminate visual feedback from scrollbars as much as possible in order to retain focus on the colour and texture images. To this end, the scrollbars were downscaled horizontally in order to hide their scroll boxes as seen for example in Figure 8.1. In this manner, it becomes impossible to tell the value for a particular dimension by looking solely at the scrollbar.

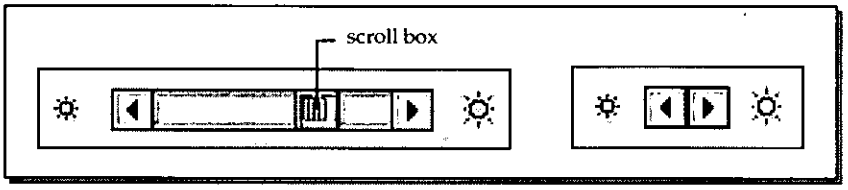


Figure 8.1: (Left) - Full scale brightness scrollbar used in the initial Sound Mosaics prototype, (Right) - The same scrollbar downscaled horizontally in order to hide the scroll box as used in the revised version of Sound Mosaics.

8.1.2 A New Visual Association for Auditory Sharpness

Another important insight gained during the first part of our evaluation studies was that there was a significant correlation between the perceived complexity of a particular sound and the perceived complexity of visual images. In more detail, as auditory sharpness increased the resulting sound was perceived as more complex and was associated by subjects with images of high complexity such as those produced by the composite texture dimension of coarseness and granularity. A closer examination of those images showed that there were two main visual characteristics of interest. First, the number of square elements was increased while their size was decreased, thus producing richer and denser textures. Second, a visual noise layer was at the same time introduced producing noise-like textures. Our observations suggested that subjects associated the first characteristic (coarseness) for sequences changing in sharpness and the second characteristic (granularity) for sounds changing in compactness. For these reasons, we decided to separate these two visual dimensions and our visualisation framework was revised in the following ways:

- Auditory sharpness was associated with texture coarseness instead of texture contrast.

- Auditory compactness was associated with texture granularity instead of the composite dimension of coarseness and granularity.

8.1.3 Towards a Single Visual Representation of Sound

The design choices discussed above had a very important consequence for the rest of the design of Sound Mosaics. In the absence of texture contrast the interaction problems caused by combining the colour and texture images within a single visual representation can be eliminated. As a result we attempted to combine the two visual representations as shown in Figure 8.2. It can be seen that the background layer is now used to represent colour information (or pitch and loudness) and the foreground layer is used to draw the texture image (timbre). It was anticipated that this design choice would significantly improve the usability of Sound Mosaics since users could focus on a single visual representation for a particular sound instead of the two visual representations incorporated in the initial Sound Mosaics prototype.

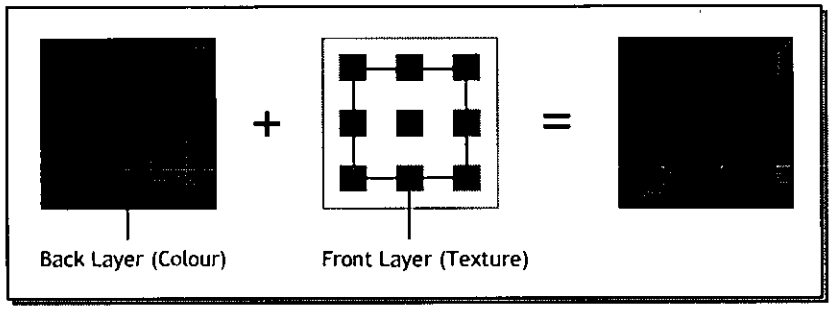


Figure 8.2: The figure shows how the colour and texture images were combined in the revised version of Sound Mosaics (see also Colour Plate E.3).

8.1.4 The Limitations of Perceptual Discrimination

As discussed in the previous chapter, perceptual discrimination for certain auditory and visual dimensions as employed in Sound Mosaics was problematic. This was particularly evident with the colour dimensions of brightness and saturation, which both employed a large number of levels. In the case of brightness, it was the large amount of pitch information that needs to be represented in a sound synthesis tool that influenced our initial design choice. However, it seems that brightness, as a sensory channel, is of inadequate capacity to represent the pitch range of musical interest. Nevertheless, the results of our empirical studies showed a strong association between brightness and pitch. In order to address this limitation without losing the *pitch-brightness* association we decided to investigate the two-dimensional theory of pitch perception as described in §3.1.1. According to this theory, pitch can be represented in terms of the dimensions of

pitch height and *tone chroma*. However, a question arises whether these dimensions are orthogonal. Although early empirical studies (e.g. (Shepard 1964)) suggested that pitch height and tone chroma are orthogonal, the results of later studies described in Deutsch (1999) indicated that there is significant interaction between the two dimensions. The reader is reminded that the two-dimensional theory of pitch perception is primarily based on the concept of *octave equivalence*, i.e. the perceptual similarity between tones whose frequencies are separated by octaves. Based on the above discussion, it seems appropriate to associate colour brightness with octave equivalence in a way that it represents octave intervals for a particular tone chroma. In this way, a significantly smaller number of brightness levels is required that is determined by the number of octaves that cover the human hearing range (approximately 10 octaves). We further suggest an association between tone chroma and colour *hue*. Although, this association is arbitrary and is not based on any empirical results, colour hue is usually used to visualise nominal data such as tone chroma (Travis 1991).

The perceptual discrimination problem encountered with saturation appears to be slightly different. Perceptual discrimination was limited for both saturation and loudness. Our own experimentation with different loudness levels suggested just-noticeable differences of 4 dB and therefore we reduced the levels for both dimensions from 100 to 23 in order to cover a loudness range of 0 - 90 dB.

8.1.5 Texture Repetitiveness and the Perceived Order Problem

The texture repetitiveness algorithm employed in the initial Sound Mosaics prototype worked by displacing an ever-increasing number of texture elements at a random distance and direction. However, the results of our evaluation studies indicated a problem with the perceived order of texture repetitiveness. In order to address this order problem we devised a new texture repetitiveness algorithm that displaces all texture elements at the same time but at an ever-increasing distance and random direction relative to the degree of sensory dissonance thus expected to produce better results in terms of the perceived order.

8.1.6 The Revised Version of Sound Mosaics

The revised implementation of Sound Mosaics (see Figure 8.3) was designed according to the design choices described in the previous sections of this chapter and the revised visualisation framework shown in Figure 8.4. It should be noted that the current Sound Mosaics prototype does not allow the selection of different hues, since further empirical work is needed in order to identify correspondences between different hues and pitch classes. However, it is anticipated that future versions of Sound Mosaics will allow users

to split the range of colour wavelengths (see Figure 4.1) into a desired number of pitch classes and use the brightness control for different octave intervals of those pitch classes.

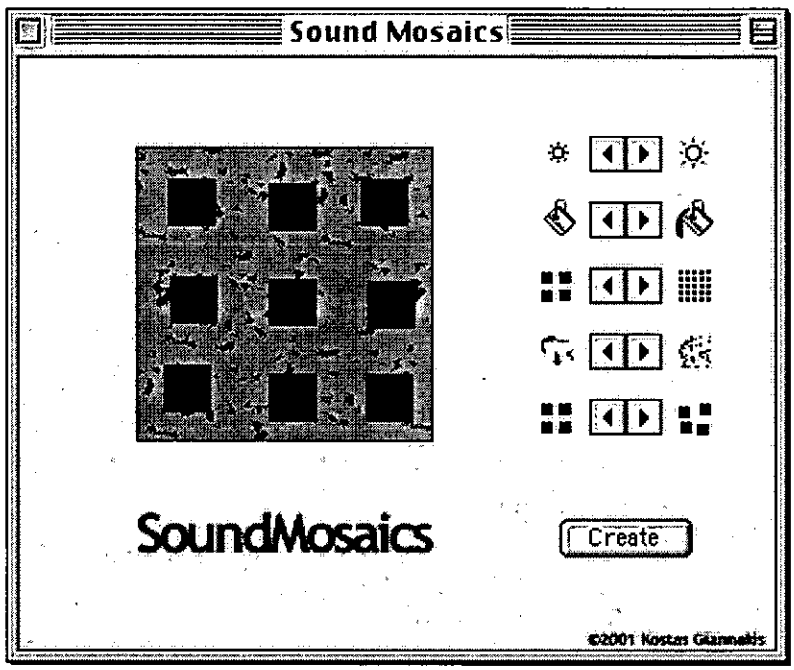


Figure 8.3: The revised Sound Mosaics prototype. In this version, the constant red hue has been arbitrarily associated with the C pitch class.

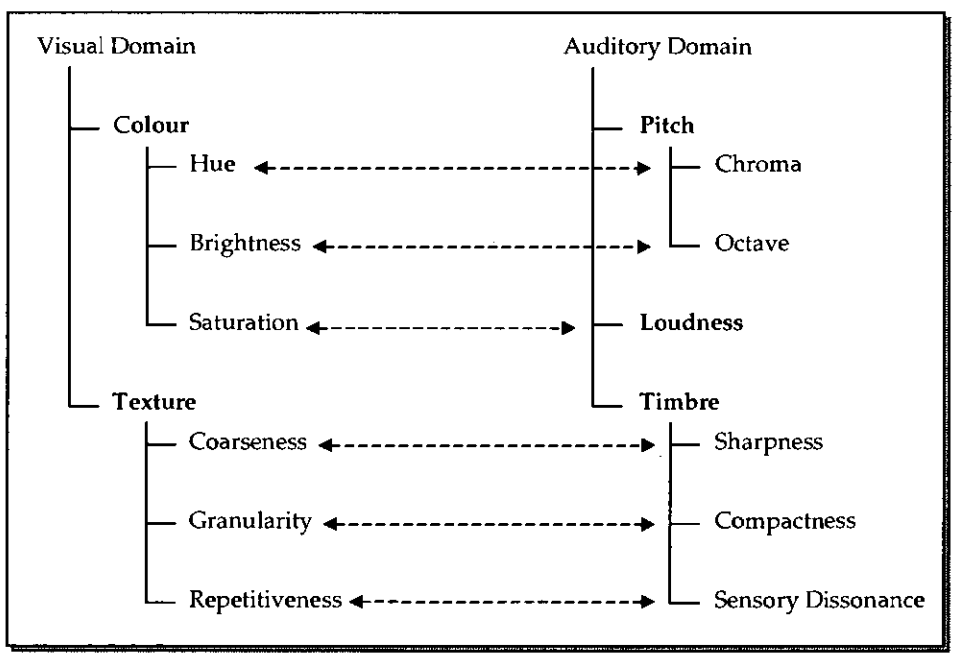


Figure 8.4: Our revised visualisation framework.

8.2 Evaluation of Sound Mosaics Revisited

The evaluation framework for the revised version of Sound Mosaics was largely based on the account given in the previous chapter and therefore only those aspects that were different in this study are presented here.

8.2.1 Challenging the Frequency-Domain Paradigm - Part II

This evaluation part was designed to compare the revised Sound Mosaics visualisation framework with the frequency-domain framework. The experimental design, tasks, apparatus and measures were the same as those described in §7.1 except for the following:

- Due to time constraints, this study was conducted with only one subject group of eight non-music subjects. Subjects were screened with the questionnaire in §A.2 in order to determine their level of musical experience and none of them had taken part in any of our previous empirical studies. We further excluded music subjects from this study because their expected familiarity with the frequency-domain framework, which could disfavour a novel framework such as Sound Mosaics.

- The Sound Mosaics stimuli used in this evaluation were produced with the revised version and processed in exactly the same way as in our previous comparison study in order to produce the frequency-domain visual stimuli. The visual stimuli for both visualisation frameworks can be found in Appendix C.

- The computer application was slightly modified to allow subjects to specify their level of confidence on-screen and not using the printed copy of the questionnaire in §A.3. However, the questions and rating scales were exactly the same.

We now present the results obtained from this experiment for each of our evaluation criteria. For simplicity of presentation the Sound Mosaics and frequency-domain frameworks are abbreviated as SM and FD respectively.

Comprehensibility

Tables 8.1 and 8.2 show the overall results for sequences that varied in either pitch or loudness for SM and FD representations respectively. In the case of SM, accuracy levels reached 50% and 81.25% for pitch height and loudness respectively. For FD

representations, accuracy levels reached 81.25% and 75% for pitch height and loudness respectively.

Sound Mosaics	Non-Music Subjects	
Selection Strategy	**Pitch Height**	**Loudness**
Brightness	**50**	18.75
Saturation	50	**81.25**
None	0	0
Total (%)	100	100

Table 8.1: Overall results for the Sound Mosaics framework obtained by non-music subjects for sequences varying in either pitch or loudness.

Frequency-Domain	Non-Music Subjects	
Selection Strategy	**Pitch Height**	**Loudness**
Brightness	18.75	75
Height	**81.25**	25
None	0	0
Total (%)	100	100

Table 8.2: Overall results for the frequency-domain framework obtained by non-music subjects for sequences varying in either pitch or loudness.

The above results give further indication that the auditory-visual associations used in SM and FD for pitch and loudness are very comprehensible although the results are slightly different from the ones obtained in our previous comparison study. Subjects in this study were more accurate in loudness sequences for both visualisation frameworks but less accurate for pitch sequences. However, the FD association between pitch and line height was again more comprehensible than the corresponding pitch-brightness association in SM.

In Tables 8.3 and 8.4 we present the overall results for sequences that varied in any of the timbral dimensions. In this case, the results are remarkably different when compared to our previous study. In more detail, the SM framework scored very high accuracy levels for all dimensions of timbre (75% for sharpness, 87.5% for compactness, and 68.75% for sensory dissonance) indicating that our revised visualisation framework has successfully resolved the problems encountered in the first comparison study. The results for FD

representations show low levels of accuracy for the dimensions of sharpness and sensory dissonance (50% and 31.25% respectively). Although the results are better for compactness sequences (81.25%) they are lower than the corresponding SM results (87.5%). These results suggest that the SM framework was more comprehensible than the FD for all dimensions of timbre.

Sound Mosaics	Non-Music Subjects		
Selection Strategy	**Sharpness**	**Comp/ness**	**S. Dissonance**
Coarseness	75	6.25	18.75
Granularity	6.25	87.50	6.25
Repetitiveness	18.75	6.25	68.75
Mixed	0	0	6.25
None	0	0	0
Total (%)	100	100	100

Table 8.3: Overall results for the Sound Mosaics framework obtained by non-music subjects for sequences varying in any of the dimensions of timbre.

Frequency-Domain	Non-Music Subjects		
Selection Strategy	**Sharpness**	**Comp/ness**	**S. Dissonance**
Line addition	50	6.25	31.25
Pixelation	12.50	81.25	31.25
Density	37.50	0	31.25
Mixed	0	12.50	6.25
None	0	0	0
Total (%)	100	100	100

Table 8.4: Overall results for the frequency-domain framework obtained by non-music subjects for sequences varying in any of the dimensions of timbre.

Intuitiveness

Figures 8.5 and 8.6 show mean completion times and confidence levels per task as obtained by non-music subjects for both visualisation frameworks. The results show that subjects were on average faster for SM representations in 5/10 cases (stimuli 1, 4, 5, 7, 9) compared to only one case where FD was faster (stimulus 3) and 4/10 cases where the two frameworks appear to be equal. In addition, subjects felt on average more confident creating sequences with SM than FD visual representations in 4/10 cases (stimuli 2, 5, 9,

10) although in the remaining cases, confidence levels were in general high and approximately equal for both visualisation frameworks.

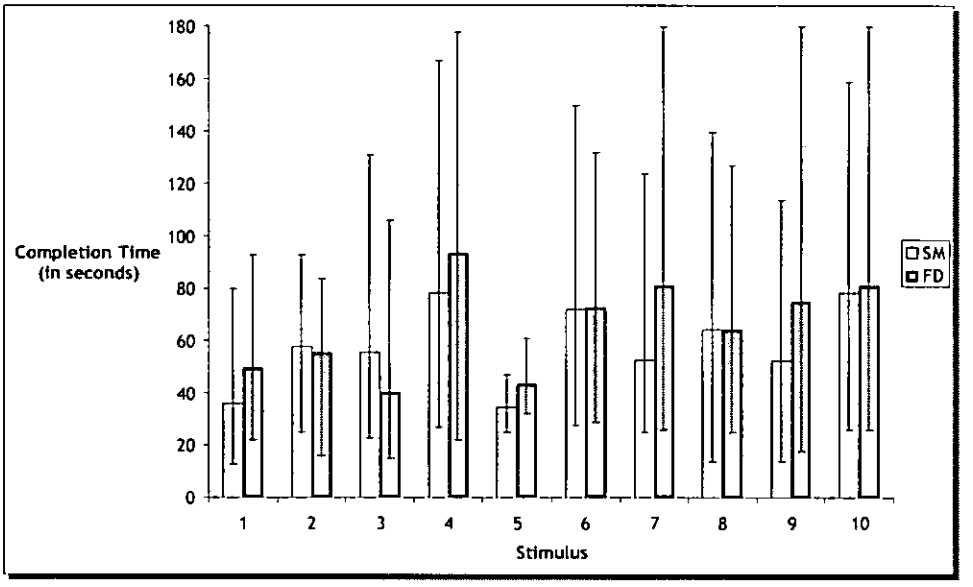


Figure 8.5: Maximum, minimum, and mean completion times per task obtained by non-music subjects for both visualisation frameworks.

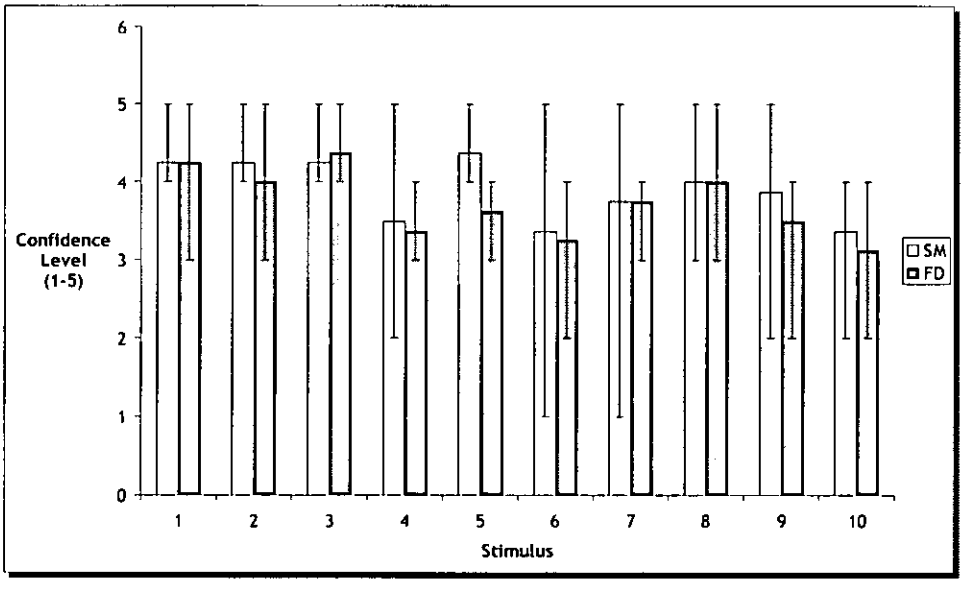


Figure 8.6: Maximum, minimum, and mean confidence levels per task obtained by non-music subjects for both visualisation frameworks.

Based on the above analysis it can be argued that the SM visualisation framework is more comprehensible and intuitive than FD visual representations of sound. Although this study was carried out with non-music subjects it is anticipated that subjects with musical experience would have performed equally well in terms of our evaluation criteria.

8.2.2 Usability Evaluation of Sound Mosaics - Part II

In this section we present a usability evaluation study of the revised version of Sound Mosaics. The method, experimental design and tasks were the same as described in §7.2 except for the following:

- We had music and non-music subject groups, each of five subjects screened with the questionnaire in §A.2. One music subject had previously taken part in the empirical study described in §4.2.

- The 19 default-target stimuli sets were designed with the revised version of Sound Mosaics.

- The Sound Mosaics prototype was again modified (see Figure 8.7) to facilitate the experimental task although in this case the "Reset" button was removed since it was used only once in our previous usability study.

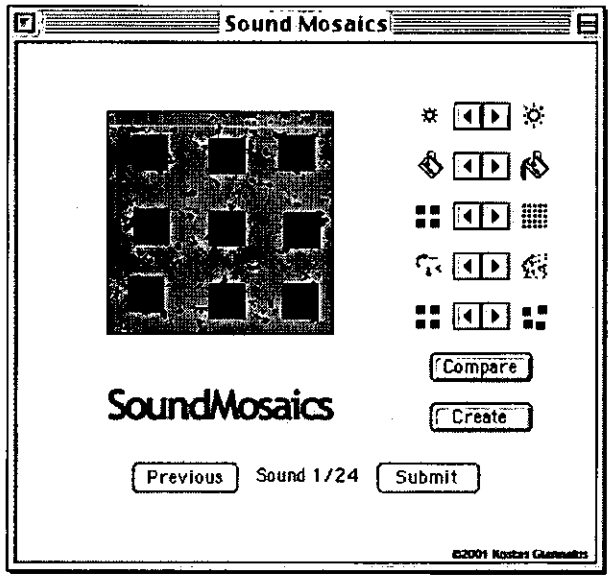


Figure 8.7: The modified Sound Mosaics prototype used in our second usability evaluation.

In the remainder of this section we present the results obtained from the above-described usability evaluation experiment for each of the usability factors under investigation.

Effectiveness

As in our previous study, the main indicator of effectiveness was the closeness of subjects' responses to the target stimulus for each task graded with the five-point scale

previously described in Table 7.10. In Tables 8.5 and 8.6, we present the scores obtained from music and non-music subjects respectively, for all 19 sound stimuli classified according to each auditory dimension. The scores are based on the average subjects' responses for each stimulus as shown in Figures 8.8 - 8.12 and as in the case of our first usability study, these figures show both music and non-music subjects' mean responses as well as the target and default levels for each stimulus (figures showing subjects' mean responses for all dimensions in each of the 19 stimuli can be found in Appendix C).

Score	L	P	S	C	SD	Total
V. High	2	2	5	4	2	15
High	0	0	0	1	2	3
Medium	0	0	0	0	0	0
Low	0	0	0	0	1	1
V. Low	0	0	0	0	0	0
Total	2	2	5	5	5	19

L: Loudness, P: Pitch, S: Sharpness,
C: Compactness, SD: S. Dissonance

Table 8.5: Overall scores for each auditory dimension obtained by music subjects.

Score	L	P	S	C	SD	Total
V. High	1	2	1	1	0	5
High	1	0	0	2	0	3
Medium	0	0	3	1	3	7
Low	0	0	1	1	1	3
V. Low	0	0	0	0	1	1
Total	2	2	5	5	5	19

L: Loudness, P: Pitch, S: Sharpness,
C: Compactness, SD: S. Dissonance

Table 8.6: Overall scores for each auditory dimension obtained by non-music subjects.

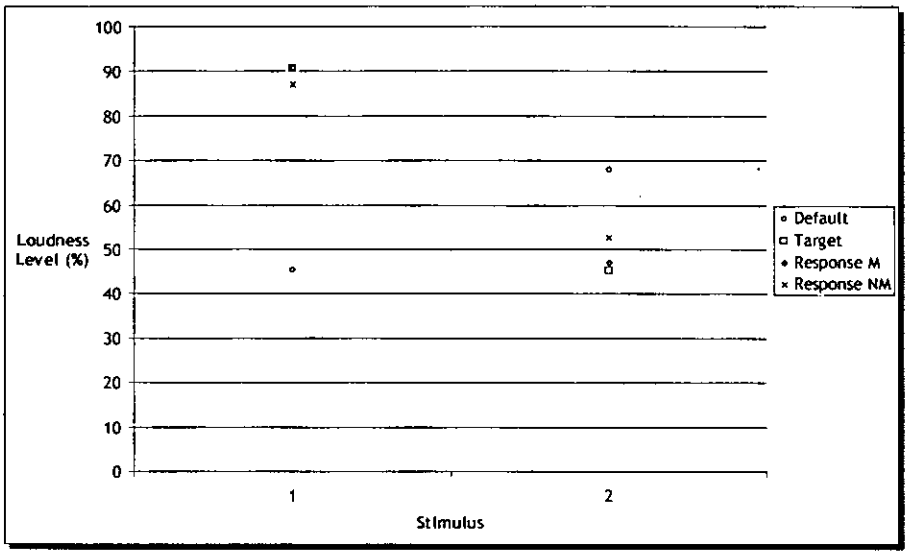


Figure 8.8: Average results obtained by both subject groups for loudness stimuli.

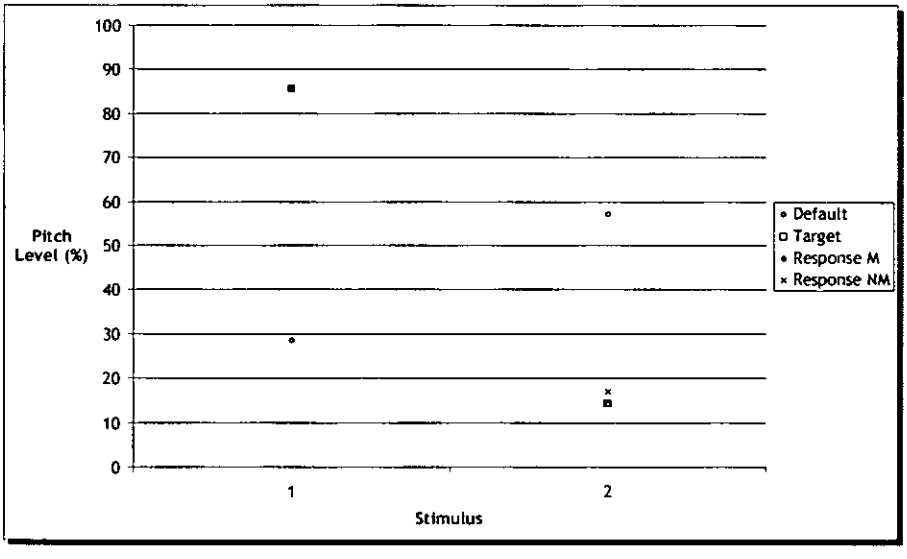


Figure 8.9: Average results obtained by both subject groups for pitch stimuli.

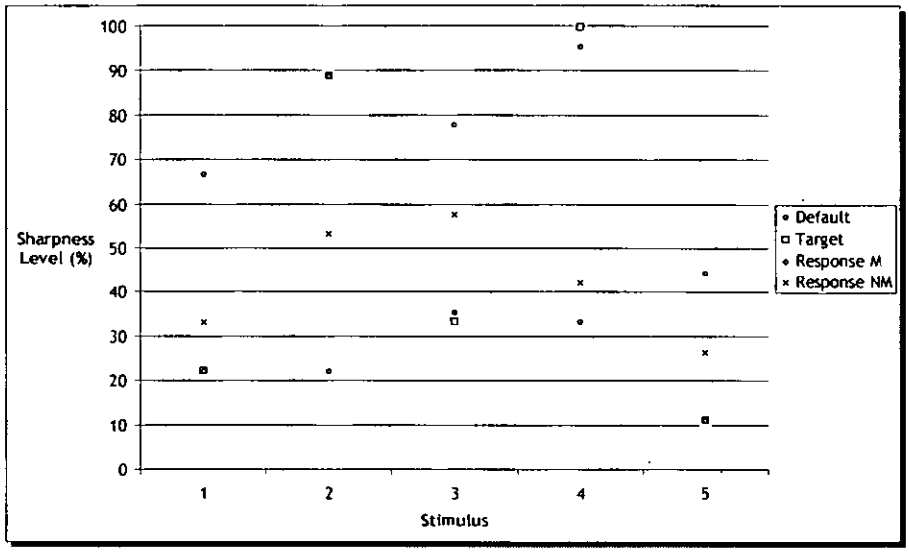


Figure 8.10: Average results obtained by both subject groups for sharpness stimuli.

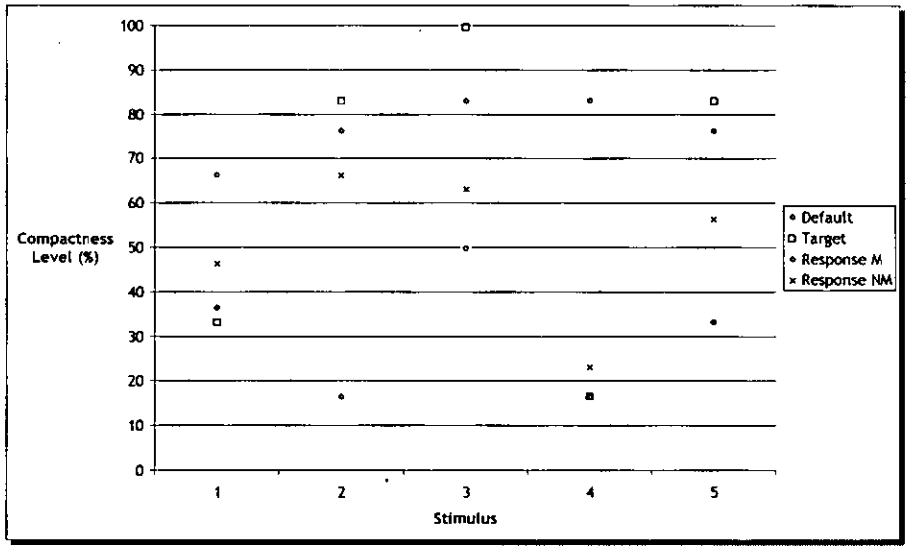


Figure 8.11: Average results obtained by both subject groups for compactness stimuli.

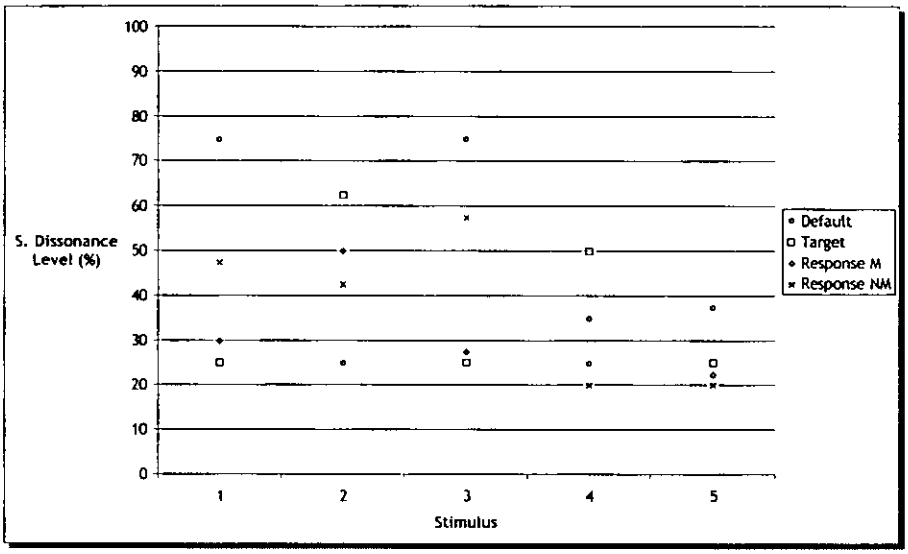


Figure 8.12: Average results obtained by both subject groups for sensory dissonance stimuli.

The sum of *high* and *very high* scores for music subjects was the same as in the first usability study (18/19 tasks) although the number of *very high* scores was larger in this study (15/19 as opposed to 13/19) indicating an improvement in subjects' performance. However, the sum of *high* and *very high* scores for non-music subjects was lower than before (8/19 compared to 14/19) with the presence of a large number of *medium* scores (7/19 tasks) especially evident for the dimensions of timbre. Nevertheless, the results are still within satisfactory levels and it is most likely that non-music subjects' performance was hampered by their unfamiliarity with the dimensions of sharpness and sensory dissonance. Non-music subjects in this study also expressed an uneasiness with the task and thought at the outset of the session that they did not expect to perform well because of their lack of musical experience. However, at the beginning of each session it was clarified to both music and non-music subjects that this study was about measuring the performance of a sound-synthesis tool and not an assessment of their musical abilities.

Efficiency

We used task completion time as the main indicator of efficiency and as before the longest time that subjects were allowed to spend on a particular stimulus was three minutes. After that period of time, the experimenter asked the subjects to submit their current choices and proceed to the next task. Figures 8.13 - 8.14 and Figures 8.15 - 8.16 show the results for music and non-music subjects respectively.

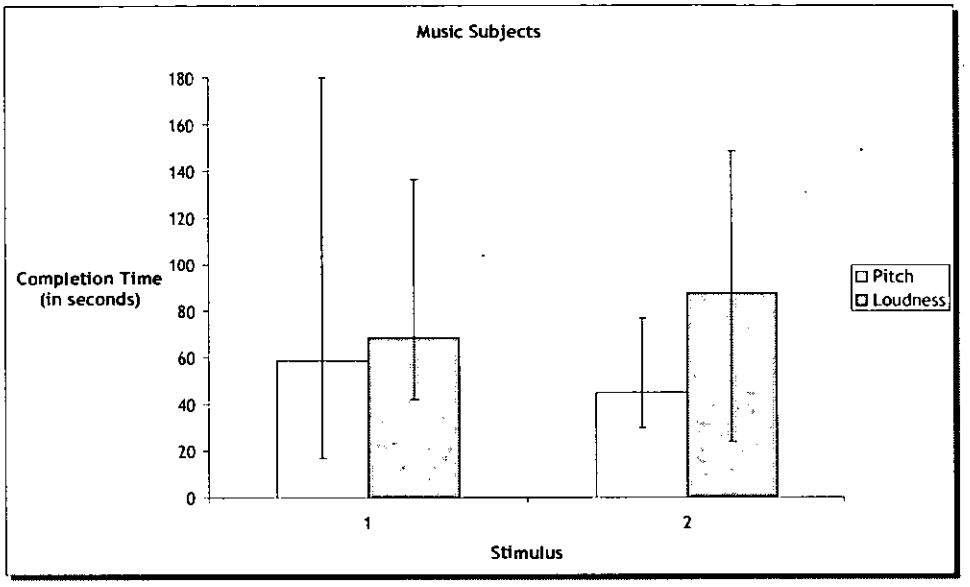


Figure 8.13: Maximum, minimum, and mean completion times obtained by music subjects for pitch and loudness stimuli.

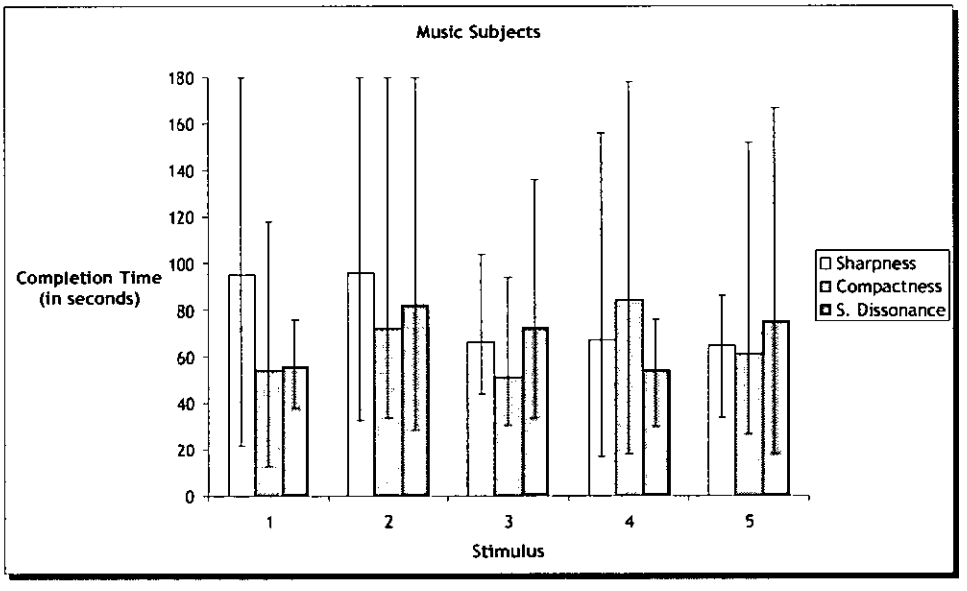


Figure 8.14: Maximum, minimum, and mean completion times obtained by music subjects for timbre stimuli.

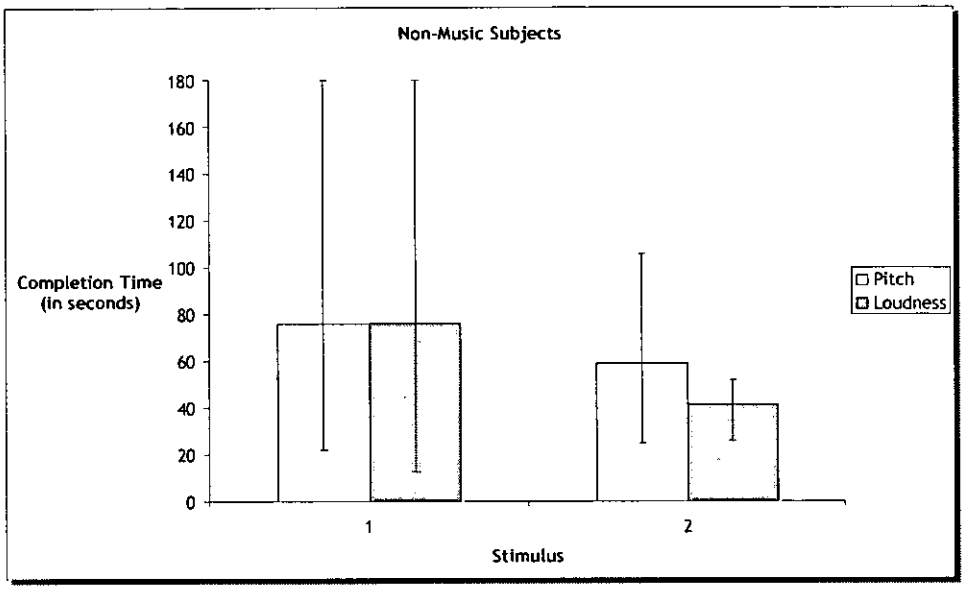


Figure 8.15: Maximum, minimum, and mean completion times obtained by non-music subjects for pitch and loudness stimuli.

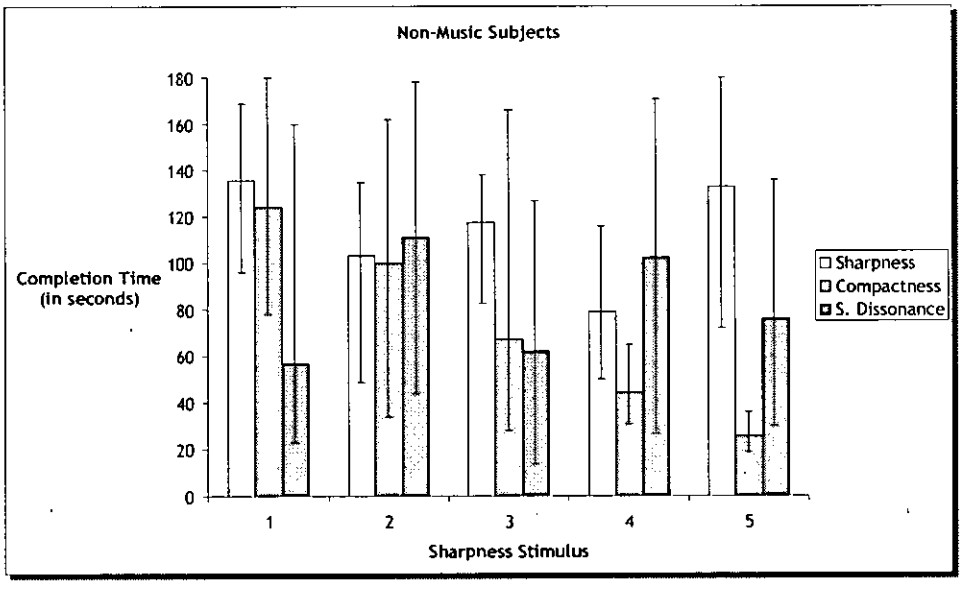


Figure 8.16: Maximum, minimum, and mean completion times obtained by non-music subjects for timbre stimuli.

In the case of music subjects, completion times for loudness and pitch stimuli were on average around the 1-minute target or less for all stimuli although the second loudness stimulus was longer than the first one. A clear decrease in completion times can be noticed for sharpness stimuli with average times very close to the 1-minute target. In the case of compactness stimuli, no decreasing trend can be observed although completion

times were around the target for 4/5 stimuli. For sensory dissonance stimuli it appears that the fifth stimulus was problematic although there is a clear decreasing trend for stimuli 2, 3, and 4. We suspect that the minor deviations in the above results were caused by the interaction between certain auditory dimensions, which added to the complexity of the particular stimuli.

Mean completion times for non-music subjects were fast for both pitch and loudness stimuli confirming that these two auditory dimensions were the easier to use. However, there is a clear difference between non-music and music subjects for the dimensions of timbre. In more detail, completion times for sharpness stimuli were clearly longer than the times obtained by music subjects. Nevertheless, there is a decreasing trend for the first four stimuli reaching times around 80 seconds although the fifth stimulus was problematic. The results are not so clear for sensory dissonance stimuli and it appears that non-music subjects had difficulty with this auditory dimension. Finally, there is clear decreasing trend for compactness stimuli achieving very fast completion times (<30 seconds).

Subjective Satisfaction

The scores obtained by music and non-music subjects using the SUS questionnaire were 72.5% and 46.5% respectively and the scores for each subject are presented in Table 8.7. If we compare these scores to the ones obtained in our previous study we can observe a 3.5% increase in music subjects' satisfaction with Sound Mosaics and a 20.5% decrease in non-music subjects' satisfaction. The latter observation gives further evidence about non-music subjects' uneasiness with the experimental task. Based on the efficiency results described earlier it can be further argued that it was the dimensions of sharpness and sensory dissonance that mainly contributed to non-music subjects' difficulty with Sound Mosaics. However, the results were very satisfactory for the intended user group of Sound Mosaics.

SUS Results	Music	Non-Music
Subject 1	60	30
Subject 2	72.5	65
Subject 3	80	55
Subject 4	77.5	32.5
Subject 5	72.5	50
Average (%)	72.5	46.5

Table 8.7: Subjective satisfaction scores obtained by music and non-music subjects for the revised Sound Mosaics prototype.

8.3 Conclusion

This chapter presented a revised implementation of Sound Mosaics based on the insights gained from the evaluation of the initial implementation described in the previous chapter. In addition, a second formative evaluation study of the revised Sound Mosaics prototype was also discussed.

As in the case of the initial prototype, the evaluation consisted of two parts. The first part dealt with a comparison study of the Sound Mosaics and frequency-domain visualisation frameworks along the criteria of comprehensibility and intuitiveness. In terms of comprehensibility, both frameworks performed satisfactorily for the dimensions of pitch and loudness and the obtained results are to a large extent consistent with our previous comparison study. However, in the case of timbre, there is clear difference between the two frameworks and the results indicated that the revised Sound Mosaics visualisation framework was more comprehensible than frequency-domain representations of sound. As far as the criterion of intuitiveness is concerned, response times and confidence levels obtained by the participants in our experiment indicated that Sound Mosaics representations were again more intuitive than frequency-domain representations. Note should be made that our second comparison study was conducted with only one subject group of non-musicians.

The second part of our evaluation was a usability study of the initial Sound Mosaics prototype. This study was conducted with both musicians and non-musicians. In the case of subjects with musical background, Sound Mosaics scored very well in terms of all three usability factors (effectiveness, efficiency, and subjective satisfaction) and improved on our previous results. However, non-musicians yielded less satisfactory results that can be primarily attributed to the subjects' unfamiliarity with the task of sound synthesis.

Based on the above, it can be argued that our revised visualisation framework and design choices improved on the weaknesses of Sound Mosaics identified in our previous evaluation studies.

9

Conclusions

In this final chapter, we present a summary of our research efforts together with the main conclusions drawn in the previous chapters of this thesis. In addition, the contributions of our work are also discussed. Finally, a critique of the research is presented, and a number of suggestions are made for further work leading from the research described in this thesis.

9.1 Summary of Thesis

The principal goal of our research was to investigate the issues related to the design of cognitively useful GUIs for computer-based sound synthesis tools. In more detail, a GUI can be thought of as consisting of a number of visual representations that bear a relationship to other elements in an application domain, in this case sound synthesis. Our review of previous and related work, as presented in Chapter 2, identified two situations of concern:

- First, current visual representations of sound are based on low-level characteristics of sound, which bear no direct relationship to perceptual experiences.

- Second, the associations between visual and auditory dimensions, i.e. the visualisation framework underlying those representations have not been empirically supported and validated.

As a result of these limitations, current sound synthesis tools require great expertise and the focus of users has shifted from the high-level musical task of sound design to the low-level and cumbersome process of understanding and controlling the visualisation framework idiomatic to each representation.

From a methodological point of view, our research attempted to address these limitations through the empirical investigation of cognitive associations between auditory and visual dimensions that led to the design of an experimental GUI for sound synthesis that allows users to specify and manipulate a set of perceptually salient auditory dimensions through the direct manipulation of visual representations. Our approach can be outlined as follows:

- A review of auditory perception studies (see Chapter 3) assisted us in the formulation of *what* aspects of sound we needed to incorporate in our investigations that are valid and prominent from a perceptual point of view and not based on physical characteristics of sound (although of course, related to them).

- Further reviews of visual perception studies (see §4.1 and §5.1) contributed in the investigation of visual dimensions that take advantage of our sensory mechanisms, such as colour and visual texture.

- The above reviews resulted in two sets of perceptual dimensions, one for each sensory domain. In order to investigate the associations (if any) between dimensions drawn from these two sets, we performed a series of empirical investigations that identified a number of important auditory-visual associations.

- In many respects, the findings of our empirical investigations formed the basis for the design of *Sound Mosaics*, a novel GUI for computer-based sound synthesis. The Sound Mosaics visualisation framework is not an arbitrary construct, but is based on the associations identified in our experiments.

- The design and implementation of Sound Mosaics followed an iterative design process, where the implementation and evaluation of an initial prototype identified various weaknesses that we attempted to address in a revised implementation.

- The success of Sound Mosaics was confirmed by comparison with the widely used frequency-domain visualisation framework. The obtained results suggested that representations created with Sound Mosaics were more comprehensible and intuitive than frequency-domain representations. In this respect our research was successful, in that it showed how an empirically derived framework for the visualisation of perceptually prominent auditory information could overcome the limitations of arbitrary auditory-visual mappings that are based on low-level characteristics of sound.

- Finally, a usability evaluation of Sound Mosaics yielded very satisfactory scores along three usability factors, namely effectiveness, efficiency, and subjective satisfaction. This was a second success for Sound Mosaics, in that the current implementation has been demonstrated to work and that there are clear potential gains leading from our approach.

9.2 Contributions

The scientific contributions of our research can be summarised as follows:

- Sound Mosaics is based completely on perceptual rather than physical dimensions of sound and has been shown to perform better than current ways of visualising sound. This is (to our knowledge) the first graphic sound synthesis tool that incorporates a perceptual model of auditory perception. The most important aspect of this model is the incorporation of dimensions of timbre.

- The empirical investigations of auditory-visual associations (see Chapters 4 and 5) were performed for the first time in the area of computer-based sound synthesis. It can be argued that our results extend those of similar approaches (e.g. investigations of synaesthesia and cross-modal associations), and our experimental methodology (e.g. the use of a three-dimensional colour model) has improved on previous efforts. In addition, the proposed auditory-visual association between loudness and saturation might be further investigated in these areas with synaesthete subjects.

- The adopted iterative design process demonstrated the importance of including prototyping and evaluation stages early in the design process, an issue that has been ignored in previous approaches to the design of computer-based sound synthesis tools. No evaluation studies of current tools exist that can demonstrate the success of the design choices. Our research is an example of how the design of computer-based sound synthesis tools may benefit form more structured approaches used in areas such as HCI. Furthermore, our research is the first that attempts to evaluate existing visual representations of sound and various strengths and weaknesses of current approaches have been identified. In particular, the evaluation of the frequency-domain (*sonogram*) representations has highlighted their inappropriateness as comprehensible and intuitive representations of sound when considering dimensions of timbre.

- Finally, the proposed visualisation framework could also be of benefit in research areas other than computer-based sound synthesis. In a similar theme, it can be used in existing approaches for music visualisation that function primarily at the macro-compositional level and lack access to the micro-compositional level (e.g. the *MidiVisualiser* (Graves *et al* 1999)). Auditory-visual mappings are also important in the development of interfaces for visually impaired users, where the goal is to communicate information that is graphical in nature and thus non-accessible to them (see Edwards (1989), Alty and Rigas (1998)). Furthermore, on-going research in the development of visual aids for visually impaired users has also focused on the direct translation of real-world images to sound (e.g. Meijer (1992),

Cronly-Dillon *et al* (2000)). For example, Meijer (1992) uses an image to sound mapping that is similar to sonogram representations of sound and forms the core of a wearable vision substitution device for the blind.

9.3 Limitations

One limitation of our empirical investigations is the limited statistical significance that we can place on the obtained results. This is a consequence of having small sample sizes that do not allow us to perform more in-depth statistical analysis and generalise the results to a larger population. However, the lack of empirical studies of auditory-visual associations and the limitations of existing approaches in this area led us at this research stage to focus more on the exploration of associations between a large number of auditory and visual dimensions and less on the statistical significance of our results. To this end, we chose to perform more empirical investigations (with a small number of subjects in each case) within the time limits of a doctoral research project. The reader should thus be aware that the conclusions that can be drawn from our empirical studies are limited.

Another limitation related to our experimental methodology is the adopted way of classifying subjects as music and non-music. Our method was to assess the musical experience of the participants in our experiments through a short questionnaire regarding subjects' experience with both traditional and computer music. A problem with this approach is that the assessment of musical experience is not objective, in the sense that it is heavily based on the subjects' own assessment of their musical abilities. Furthermore, our musical experience test was not designed to test subjects' perceptual abilities (e.g. pitch perception), which may have affected their performance during the experiments. Recent studies (e.g. Edwards *et al* (2000)) have argued that the assessment of subjects' musical ability is an issue that deserves more careful treatment and have pointed out the need for a standard test of musical ability that could be incorporated in evaluation experiments. The Musical Aptitude Test (MAT) (see Hakinson *et al* (1999), Edwards *et al* (2000)) is a promising approach towards that direction and although it is primarily targeted for the evaluation of auditory interfaces it could also be beneficial in our research and related areas where an assessment of the subjects' musical ability is an integral part.

The models of auditory and visual perception are of course only partial, since various important perceptual dimensions were not investigated. This was also due to time constraints and the lack of perceptually based analyses of those dimensions (e.g. shape). Although, the scope of our models is limited, it is at the same time open-ended allowing further dimensions to be added through further work. Finally, it should be further noted

that these models are not perceptually uniform in the sense that equal steps in any of the dimensions incorporated in Sound Mosaics do not necessarily correspond to equally perceived changes.

9.4 Methodology

From a methodological point of view, our research has been largely based on an empirical method for the identification of associations between auditory and visual dimensions. Our goal has been to overcome the limitations that might be presented by arbitrary associations as discussed earlier in Chapter 2. An alternative approach could have been to directly involve user of computer-based sound synthesis tools early in the design process (e.g. formal interviews). However, the novelty of this interface and therefore the lack of experienced users makes the above approach inappropriate. In addition, auditory-visual associations that arise from users' personal experiences may be completely different for each user thus making it difficult to establish associations that will hold for a large number of users without the need of learning them. For these reasons, the adopted research methodology based on observation studies under controlled conditions seems appropriate for our research goals.

9.5 Further Work

In this section, we present suggestions for further work, leading from the research described in this thesis.

One of our short-term goals is to perform more rigorous empirical investigations of the proposed auditory-visual associations. In more detail, we plan to incorporate larger subjects groups than the ones used in this thesis and use statistical methods that will strengthen the reliability and validity of our conclusions. At the same time, a series of small-scale experiments could be made in order to investigate additional correspondences between auditory and visual dimensions. For example, our association between hue and tone chroma needs to be further tested for empirical validation although our evaluation studies of Sound Mosaics suggested that the mapping was successful. Other candidate perceptual dimensions for these experiments are temporal and spatial characteristics in both the auditory and visual domains. Our use of simple experimental designs in this thesis has indicated that such investigations could be useful and reliable sources for design ideas and could assist us in building and testing future prototype versions of Sound Mosaics.

Our research did not attempt to incorporate perceptually uniform scales for the associations between auditory and visual dimensions. To develop such scales would

require rigorous psychophysical experiments — such experiments would be a natural extension leading from the research we described in this thesis.

With respect to implementation issues of Sound Mosaics, a number of extensions to the current implementation can be made. One desirable extension is to include an analysis stage to the Sound Mosaics mechanism in order to allow the analysis and modification of existing sounds, thus extending the synthesis capabilities of Sound Mosaics. Another extension could be the real-time control of auditory dimensions. As discussed in §6.2.1, the current implementation allows real-time control of visual dimensions only. It is envisaged that the real-time control of auditory dimensions will significantly enhance the users' interaction with Sound Mosaics and will widen the applicability of Sound Mosaics for other purposes (e.g. performing). However, it is not clear whether the selected programming environment (JAVA and Csound) can support these aims in a straightforward manner. A number of alternative development environments are being currently considered that are designed to facilitate the real-time generation of both sound and graphics within the same programming environment (e.g. *SuperCollider* (McCartney 1996, 2000), *Siren/Squeak* (Pope 1998)).

As a long-term goal, we plan to investigate a number of design ideas that could enhance the users' interaction with Sound Mosaics. In more detail, it could be more natural for users to perform some actions directly on the colour/texture panel than interacting with scrollbars. For example, the amount of texture repetitiveness could be specified by selecting and editing the desired texture elements directly on the texture image. This will require a more sophisticated way of image processing than the one currently employed in order to interpret the user's intentions. Furthermore, when an adequate set of auditory-visual associations has been gathered, the analysis of natural textures could also be considered by allowing users to import images created by other means.

9.6 Epilogue

Our age is characterised by a growing interest in the application of computers in arts. Computer technology has become an everyday tool for creative expression in almost every area of human activity. As a result, traditional ways of expression have been transformed and, most importantly, new forms of electronic art have been introduced. Music, as both a form of art and a field of science, has undergone significant changes due to the application of computers in areas such as sound analysis and synthesis, composition, performance, etc. It is now possible to probe into these different areas of musical knowledge and experiment with new ideas and tools.

We hope that the work presented in this thesis will contribute to future investigations of auditory-visual associations and the development of cognitively useful graphical user interfaces for music compositional processes and related purposes.

R

References

Alty, J. L. & Rigas, D. I. (1998) 'Communicating Graphical Information to Blind Users Using Music: The Role of Context' in *Proceedings of the 1998 CHI - ACM Conference on Human Factors in Computing Systems*, Los Angeles.

Amadasun, M. & King, R. (1989) 'Textural features corresponding to textural properties' *IEEE Transactions on Systems, Man, and Cybernetics* Vol. 19 No. 5, pp 1264-1274.

Arnheim, R. (1969) *Visual thinking*, University of California Press, Berkeley CA.

Arnheim, R. (1974) *Art and visual perception: A psychology of the creative eye*, University of California Press, Berkeley CA.

ASA (1960) *Acoustical Terminology SI, 1-1960*, American Standards Association, New York.

Barrass, S. (1997) *Auditory Information Design*, PhD Thesis, The Australian National University (unpublished).

Bianchini, R. & Cipriani, A. (2000) *Virtual sound*, Contempo, Rome.

Bimbo, A. del (1999) *Visual information retrieval*, Morgan Kaufmann Publishers, San Francisco CA.

Bismarck, G. von. (1974a) 'Timbre of steady sounds: A factorial investigation of its verbal attributes' *Acustica* Vol. 30, pp 146-159.

Bismarck, G. von. (1974b) 'Sharpness as an attribute of the timbre of steady sounds' *Acustica* Vol. 30, pp 159-172.

Boulanger, R. C. (2000) *The Csound book*, MIT Press, Cambridge MA.

Bregman, A. S. (1990) *Auditory scene analysis: The perceptual organization of sound*, MIT Press, Cambridge MA.

Brooke, J. (1996) 'SUS: A 'quick and dirty' usability scale' in P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (eds) *Usability Evaluation in Industry*, Taylor and Francis, London

Butler, D. (1992) *The musician's guide to perception and cognition*, Schirmer Books, New York.

Caivano, J. L. (1994) 'Color and sound: Physical and psychophysical relations' *Color Research and Application* Vol. 1 No. 2, pp 126-132.

Cronly-Dillon, J., Persaud, K. & Blore, R. (2000) 'Blind subjects construct conscious mental images of visual scenes encoded in musical form' *Proceedings of the Royal Society of London Series B - Biological Sciences* Vol. 267, pp 2231-2238.

Cytowic, R. (1993) *The man who tasted shapes,* Abacus, London.

Dann, K. T. (1998) *Bright colors falsely seen: Synaesthesia and the search for transcendental knowledge,* Yale University Press, New Haven and London.

Deutsch, D. (1999) 'The Processing of Pitch Combinations' in D. Deutsch (ed) *The Psychology of Music,* Academic Press, San Diego CA.

Digidesign (1995) *Turbosynth,* Computer Application, Digidesign Inc.

Dixon Ward, W. (1999) 'Absolute Pitch' in D. Deutsch (ed) *The Psychology of Music,* Academic Press, San Diego CA.

Dodge, C. & Jerse, T. A. (1997) *Computer music: Synthesis, composition, and performance,* Schirmer Books, New York.

Eckel, G. (1992) 'Manipulation of Sound Signals Based on Graphical Representation: A Musical Point of View' in *Proceedings of the 1992 International Workshop on Models and Representations of Musical Signals,* Capri.

Edwards, A. (1989) 'Soundtrack: An auditory interface for blind users' *Human-Computer Interaction* Vol. 4 No. 1, pp 45-66.

Edwards, A., Challis, B. P., Hankinson, J. C. K. & Pirie, F. L. (2000) 'Development of a Standard Test of Musical Ability for Participants in Auditory Interface Testing' in *Proceedings of ICAD 2000,* Atlanta.

Ehresman, D. & Wessel, D. (1978) Perception of Timbral Analogies. *IRCAM* Reports, 13/78.

Ethington, R. & Punch, B. (1994) 'Seawave: A system for musical timbre description' *Computer Music Journal* Vol. 18 No. 1, pp 30-39.

Francos, J. Meiri, A., & Porat, B. (1991) 'Modelling of the texture structural components using 2-D deterministic random fields' *Visual Communications and Image Processing* Vol. SPIE 1666, pp 554-565.

Fairchild, M. D. (1998) *Color appearance models,* Addison-Wesley Longman, Reading MA.

Fitz, K., Walker, W., & Haken, L. (1992) 'Extending the McAulay-Quatieri Analysis for Synthesis with a Limited Number of Oscillators' in *Proceedings of the International Computer Music Conference 1992,* San Jose.

Fitz, K., Haken, L., & Holloway, B. (1995) 'Lemur - A Tool for Timbre Manipulation' in *Proceedings of the International Computer Music Conference 1995,* Banff.

Fletcher, H. & Munson, W. A. (1933) 'Loudness, its definition, measurement, and calculation' *Journal of the Acoustical Society of America* Vol. 5, pp 82-108.

Foley, J. D., van Dam, A., Feiner, S. K., Hughes, J. F., & Phillips, R. L. (1994) *Introduction to computer graphics,* Addison-Wesley, Reading MA.

Foner, L. N. (1996) 'Artificial Synesthesia via Sonification: A Wearable Augmented Sensory System' in *Proceedings of ICAD 96,* Palo Alto.

Fortner, B. & Meyer, T. E. (1997) *Number by colors: A guide to using color to understand technical data,* Springer-Verlag, New York.

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000) 'Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?' in *Proceedings of the 2000 CHI - ACM Conference on Human Factors in Computing Systems,* The Hague.

Giannakis, K. & Smith, M. (1999) 'Imaging Soundscapes' in *Extended abstracts of CMI 99: The 6th conference of the International Society for Systematic and Comparative Musicology,* Oslo.

Giannakis, K. & Smith, M. (2000a) 'Towards a Theoretical Framework for Sound Synthesis based on Auditory-Visual Associations' in *Proceedings of the AISB'00 Symposium on Creative and Cultural Aspects and Applications of AI and Cognitive Science,* Birmingham.

Giannakis, K. & Smith, M. (2000b) 'Auditory-Visual Associations for Music Compositional Processes: A Survey' in *Proceedings of the 2000 International Computer Music Conference*, Berlin.

Giannakis, K. & Smith, M. (in press) 'Imaging Soundscapes: Identifying Cognitive Associations between Auditory and Visual Dimensions' in R. I. Godøy and H. Jørgensen (eds) *Musical Imagery*, Swets and Zeitlinger, Lisse.

Goldberg, T. & Schrack, G. (1986) 'Computer-aided correlation of musical and visual structures' *Leonardo* Vol. 19 No. 1, pp 11-17.

Graves, A., Hand, C., & Hugill, A. (1999) 'MidiVisualiser: Interactive music visualisation using VRML' *Organised Sound* Vol.4 No. 1, pp 15-23.

Grey, J. M. (1975) *Exploration of Musical Timbre*, PhD Thesis, Report No. STAN-M-2, Stanford University, Stanford CA.

Gupta, A., & Jain, R. (1997) 'Visual Information Retrieval' *Communications of the ACM* Vol. 40 No. 5, pp 71-79.

Hakinson, J. C. K., Challis, B. P., & Edwards, A. (1999) *MAT: A Tool for Measuring Musical Ability*, Technical Report YCS 322, University of York, Department of Computing Science.

Harrison, J. E. & Baron-Cohen, (1997) 'Synaesthesia: An Introduction' in S. Baron-Cohen & J. E. Harrison (eds) *Synaesthesia*, Blackwell Publishers, Oxford.

Healey, C. & Enns, J. (1998) 'Building Perceptual Textures to Visualize Multidimensional Datasets' in Proceedings of the *IEEE Visualization 1998*, Research Triangle Park.

Heaps, C. & Handel, S. (1999) 'Similarity and features of natural textures' *Journal of Experimental Psychology: Human Perception and Performance* Vol. 25 No. 2, pp 299-320.

Holtzman, S. R, (1994) *Digital Mantras*, MIT Press, Cambridge MA.

Hubbard, T. L. (1996) 'Synesthesia-like mappings of lightness, pitch, and melodic interval' *American Journal of Psychology* Vol. 109 No. 2, pp 219-238.

Hutchinson, W. & Knopoff, L. (1978) 'The acoustic component of western consonance' *Interface,* Vol. 7, pp 1-29.

IRCAM (1995) *Audiosculpt,* Computer Application, IRCAM (`http://www.ircam.fr/`).

ISO 9241-11 (1998) *Ergonomic requirements for office work with visual display terminals - Part 11: Guidance on usability,* International Organization for Standardization.

Jackson, R., MacDonald, L. & Freeman, K. (1994) *Computer generated colour: A practical guide to presentation and display,* John Wiley & Sons, Chichester.

Karinthi, P. Y. (1991) 'A contribution to musicalism: An attempt to interpret music in painting' *Leonardo* Vol. 24 No. 4, pp 401-405.

Kendall, R. & Carterette, E. C. (1997) 'Difference Thresholds for Timbre Related to Spectral Centroid' in *Proceedings of the ICPMC 97,* Uppsala.

LPA (1998) *MacProlog32,* Computer Application, Logic Programming Associates Limited.

Leman, M. (1993) 'Symbolic and Subsymbolic Description of Music' in G. Haus (ed) *Music Processing,* Oxford University Press, Oxford.

Lesbros, V. (1996) 'From images to sounds: A dual representation' *Computer Music Journal* Vol. 20 No. 3, pp 59-69.

Liu, F. & Picard, R. W. (1996) 'Periodicity, directionality, and randomness: Wold features for image modeling and retrieval' *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 18 No. 7, pp 722-733.

Low, A. (1991) *Introductory computer vision and image processing,* McGraw-Hill, London.

McAdams, S. (1999) 'Perspectives on the contribution of timbre to musical structure' *Computer Music Journal* Vol. 23 No. 3, pp 85-102.

McAdams, S., Winsberg, S., Donnadiey, S., De Soete, G., Krimphoff, J. (1995) 'Perceptual scaling of synthesized musical timbres. Common dimensions, specificities, and latent subject classes' *Psychological Research* Vol. 58, pp 177-192.

McCartney, J. (1996) 'SuperCollider, a New Real Time Synthesis Language' in *Proceedings of the 1996 International Computer Music Conference*, Hong Kong.

McCartney, J. (2000) 'A New, Flexible Framework for Audio and Image Synthesis' in *Proceedings of the 2000 International Computer Music Conference*, Berlin.

Malloch, S. N. (1997) *Timbre and Technology: An Analytical Partnership*, PhD Thesis, University of Edinburgh (unpublished).

Marks, L. E. (1975) 'On colored-hearing synesthesia: Cross-modal translations of sensory dimensions' *Psychological Bulletin* Vol. 82 No. 3, pp 303-331.

Martino, G. & Marks, L. E. (2000) 'Cross-modal interaction between vision and touch: The role of synesthetic correspondence' *Perception* Vol. 29, pp 745-754.

Meijer, P. B. L. (1992) 'An experimental system for auditory image representations' *IEEE Transactions on Biomedical Engineering* Vol. 39 No. 2, pp 112-121.

Miranda, E. R. (1998) *Computer sound synthesis for the electronic musician*, Focal Press, Oxford.

Miranda, E. R. (1994) *An Artificial Intelligence Approach to Sound Design*, PhD Thesis, University of Edinburgh (unpublished).

Moore, B. C. J. (1997) *An introduction to the psychology of hearing*, Academic Press, San Diego CA.

Narayanan, N. H. & Hübscher, R. (1998) 'Visual Language Theory: Towards a Human-Computer Interaction Perspective' in K. Marriott and B. Meyer (eds) *Visual Language Theory*, Springer-Verlag, New York.

Nelson, P. (1997) 'The UPIC system as an instrument of learning' *Organised Sound* Vol. 2 No. 1, pp 35-42.

Newman, W. M. & Lamming, M. G. (1995) *Interactive system design*, Addison-Wesley Publishers, Harlow.

Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E. H., Petkovic, D., Yanker, P., Faloutsos, C., & Taubin, G. (1993) 'The QBIC Project: Querying Images by

Content Using Color, Texture, and Shape' in *Proceedings of the SPIE Conference on storage and retrieval for image and video databases*, San Jose.

Nielsen, J. (1993) *Usability engineering*, Academic Press, London.

Noyes, J. & Baber, C. (1999) *User-centred design of systems*, Springer-Verlag, London.

Olson, H. F. (1967) *Music, physics and engineering*, Dover Publications, New York.

Padgham, C. (1986) 'The scaling of the timbre of the piping organ' *Acustica* Vol. 60, pp 189-204.

Paterno', F. (2000) *Model-based design and evaluation of interactive applications*, Springer-Verlag, London.

Peacock, K. (1988) 'Instruments to perform color-music: Two centuries of technological experimentation' *Leonardo* Vol. 21 No. 4, pp 397-406.

Pierce, J. (1999) 'The Nature of Musical Sound' in D. Deutsch (ed) *The Psychology of Music*, Academic Press, San Diego CA.

Plomp, R. (1976) *Aspects of tone sensation*, Academic Press, New York.

Plomp, R. & Levelt, W. J. M. (1965) 'Tonal consonances and critical bandwidth' *Journal of the Acoustical Society of America* Vol. 38, pp 548-560.

Pocock-Williams, L. (1992) 'Toward the automatic generation of visual music' *Leonardo* Vol. 25 No 1, pp 445-452.

Pollard, H. F. (1982) 'A tristimulus method for the specification of musical timbre' *Acustica* Vol. 51, pp 162-171.

Pope, S. T. (1998) 'The Siren Music/Sound Package for Squeak Smalltalk' in *Proceedings of the 1998 ACM Conference on Object-Oriented Programming Systems, Languages, and Applications*, Vancouver.

Rasch, R. & Plomp, R. (1999) 'The Perception of musical Tones' in D. Deutsch (ed) *The Psychology of Music*, Academic Press, San Diego CA.

Rao, A. R. (1990) *A taxonomy for texture description and identification,* Springer-Verlag, New York.

Rao, A. R. & Lohse, G. L. (1993) 'Identifying high level features of texture perception' *Graphical Models and Image Processing* Vol. 55 No. 3, pp 218-233.

Rao, A. R. & Lohse, G. L. (1996) 'Towards a texture naming system: Identifying relevant dimensions of texture' *Vision Research* Vol. 36 No. 11, pp 1649-1669.

Reeves, B., & Nass, C. (2000) 'Perceptual Bandwidth' *Communications of the ACM* Vol. 43 No. 3, pp 65-70.

Risset, J. & Wessel, D. (1999) 'Exploration of Timbre by Analysis and Synthesis' in D. Deutsch (ed) *The Psychology of Music,* Academic Press, San Diego CA.

Roads, C. (1996) *The computer music tutorial,* MIT Press, Cambridge MA.

Rogowitz, B. E. & Treinish, L. A. (1998) 'Data visualization: The end of the rainbow' *IEEE Spectrum* Vol. 12, pp 52-59.

Russell, P. (1995) *PowerSynthesiser,* Computer Application, University of Sussex.

Schoffer, N. (1985) 'Sonic and visual structures: Theory and experiment' *Leonardo* Vol. 18 No. 2, pp 59-68.

Sebba, R. (1991) 'Structural correspondence between music and color' *Color Research and Application* Vol. 16 No. 2, pp 81-88.

Serra, X. (1997a) 'Current perspectives in the digital synthesis of musical sounds', *Formats* No. 1, Pompeu Fabra University, Barcelona.

Serra, X. (1997b) 'Musical Sound Modelling with Sinusoids plus Noise' in C. Roads, S. Pope, A. Piccialli, and G. de Poli (eds) *Music Signal Processing,* Swets & Zeitlinger, Lisse.

Sethares, W. A. (1999) *Tuning, timbre, spectrum, scale,* Springer-Verlag, London.

Shepard, R. (1999) 'Pitch Perception and Measurement' in P. R. Cook (ed) *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics,* MIT Press, Cambridge MA.

Shneiderman, B. (1998) Designing the user interface: Strategies for effective human-computer interaction, Addison-Wesley Longman, Reading MA.

Slawson, W. (1985) *Sound color*, University of California Press, Berkeley CA.

Sloboda, J. A. (1985) *The musical mind: The cognitive psychology of music*, Oxford University Press, Oxford.

Smith, A. R. (1978) 'Color gamut transform pairs' *Computer Graphics* Vol. 12, pp 12-19.

Stevens, S. S. (1936) 'A scale for the measurement of a psychological magnitude', *Psychological Review* Vol. 43, pp 405-416.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937) 'A scale for the measuremnt of the psychological magnitude of pitch' *Journal of the Acoustical Society of America* Vol. 8, pp 185-190.

Tamura, H., Mori, S., & Yamawaki, T. (1978) 'Textural features corresponding to visual perception' *IEEE Transactions on Systems, Man, and Cybernetics* Vol. 8 No. 6, pp 460-473.

Taylor, C. (2000) 'The Physics of Sound' in P. Kruth and H. Stobart (eds) *Sound*, Cambridge University Press, Cambridge.

Tomita, F. & Tsuji, S. (1990) *Computer analysis of visual textures*, Kluwer Academic Publishers, Norwell MA.

Travis, D. (1991) *Effective color displays: Theory and practice*, Academic Press, London.

Tsai, J. L. & Perng, D. B. (1998) 'An intuitive and device-independent method of generating color atlases for electronic displays' *The Visual Computer* Vol. 14, pp 328-342.

U & I Software (1998) *MetaSynth*, Computer Application, U & I Software (`http://www.uisoftware.com/`).

Vercoe, B. (1993) *Csound*, Computer Application, Massachusetts Institute of Technology.

Vertegaal, R. & Eaglestone, B. (1998) 'Looking for Sound? Selling Perceptual Space in Hierarchically Nested Boxes' in *Proceedings of the 1998 CHI - ACM Conference on Human Factors in Computing Systems*, Los Angeles.

Walters, J. L. (1997) 'Sound, code, image' *EYE* Vol. 7 No. 26, pp 24-35.

Ware, C. (2000) *Information visualization: Perception for design*, Academic Press, San Diego CA.

Ware, C. & Knight, W. (1992) 'Orderable Dimensions of Visual Texture for Data Display: Orientation, Size, and Contrast' in *Proceedings of the 1992 CHI - ACM Conference on Human Factors in Computing Systems*, Monterey.

Ware, C. & Knight, W. (1995) 'Using visual texture for information display' *ACM Transactions on Graphics* Vol. 14 No. 1, pp 3-20.

Wells, A. (1980) 'Music and visual color: A proposed correlation' *Leonardo* Vol. 13 No. 1, pp 101-107.

Whitney, J. H. (1980) *Digital harmony*, Byte Books, Peterborough NH.

Whitney, J. H. (1991) 'Fifty years of composing computer music and graphics: How time's new solid state tractability has changed audio-visual perspectives' *Leonardo* Vol. 24 No. 5, pp 597-599.

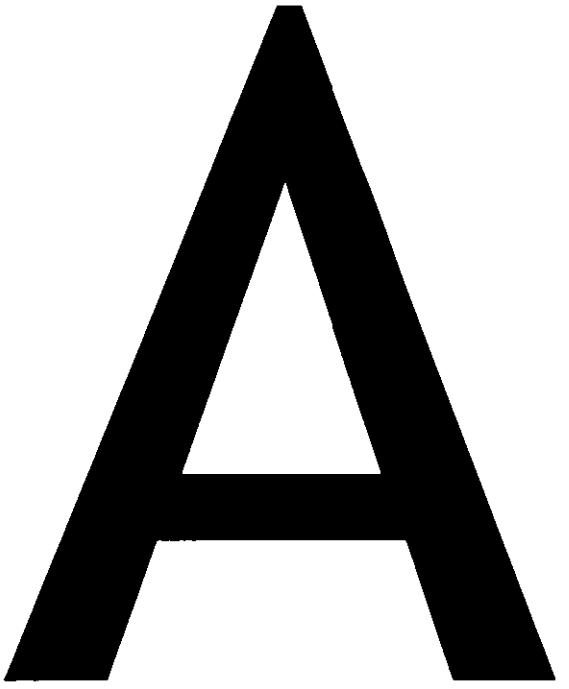
Wishart, T. (1996) *On sonic art*, Revised edition, Harwood Academic Publishers, Amsterdam.

De Witt, T. (1987) 'Visual music: Searching for an aesthetic' *Leonardo* Vol. 20 No. 2, pp 115-122.

Xenakis, I. (1992) *Formalized music*, Pendragon Press, New York.

Zwicker, E. & Fastl, H. (1999) *Psychoacoustics: Facts and models*, Springer-Verlag, Berlin.

Appendix



A.1 Questionnaire I

The questionnaire below is reprinted as used in the empirical investigations described in Chapters 4 and 5.

Background Information

Dear Colleague,

Thank you for agreeing to participate in this experiment. The main goal of this research is to investigate more intuitive ways of representing musical sound in computers. The results of this experiment will be deployed in academic research including conference papers and journal articles.

Please note that your personal details will be treated with confidence.

Name: _____

Sex: Female ☐ Male ☐

Age: ☐ years

Music Education:

None ☐ Basic Music Theory ☐ Advanced Studies ☐ Other ☐

Please specify: _____

Do you play any musical instruments?

Yes ☐ No ☐

Please specify: _____

Experience with computer-based music applications:

None ☐ Little ☐ Average ☐ Great ☐

Please specify: _____

A.2 Questionnaire II

The questionnaire below is reprinted as used in the empirical investigations described in Chapters 7 and 8.

Background Information

Dear Colleague,

Thank you for agreeing to participate in this experiment. The main goal of this research is to investigate more intuitive ways of representing musical sound in computers. The results of this experiment will be deployed in academic research including conference papers and journal articles.

Please note that your personal details will be treated with confidence.

Name: _____

Sex: Female ☐ Male ☐

Age: ☐ years

Music Education:

None ☐ Basic Music Theory ☐ Advanced Studies ☐ Other ☐

Please specify:_____

Do you play any musical instruments?

Yes ☐ No ☐

Please specify:_____

Experience with computer-based music applications:

None ☐ Little ☐ Average ☐ Great ☐

Please specify:_____

Experience with the following computer-based sound synthesis applications:

UPIC: None ☐ Little ☐ Average ☐ Great ☐

PHONOGRAMME: None ☐ Little ☐ Average ☐ Great ☐

METASYNTH: None ☐ Little ☐ Average ☐ Great ☐

A.3 Questionnaire III

The questionnaire below is reprinted as used to assess confidence levels in the empirical investigation described in Chapter 7.

Please indicate the degree to which you agree or disagree with the following statement:

"I feel very confident that there is a very good match between the sequence of images I have just created and the current sound sequence."

	Strongly Disagree				Strongly Agree
Sound sequence #1:	1	2	3	4	5
Sound sequence #2:	1	2	3	4	5
Sound sequence #3:	1	2	3	4	5
Sound sequence #4:	1	2	3	4	5
Sound sequence #5:	1	2	3	4	5
Sound sequence #6:	1	2	3	4	5
Sound sequence #7:	1	2	3	4	5
Sound sequence #8:	1	2	3	4	5
Sound sequence #9:	1	2	3	4	5
Sound sequence #10:	1	2	3	4	5
Sound sequence #11:	1	2	3	4	5
Sound sequence #12:	1	2	3	4	5
Sound sequence #13:	1	2	3	4	5
Sound sequence #14:	1	2	3	4	5
Sound sequence #15:	1	2	3	4	5
Sound sequence #16:	1	2	3	4	5
Sound sequence #17:	1	2	3	4	5
Sound sequence #18:	1	2	3	4	5
Sound sequence #19:	1	2	3	4	5
Sound sequence #20:	1	2	3	4	5

A.4 Questionnaire IV

The questionnaire below is based on the Subjective User Satisfaction (SUS) questionnaire (Brooke 1996) as used to assess users' subjective satisfaction with Sound Mosaics during the usability evaluation studies described in Chapters 7 and 8.

Please indicate the degree to which you agree or disagree with each of the statements below. Your responses should be as immediate as possible. If you feel that you cannot respond to a particular statement circle the centre point (3) for that statement.

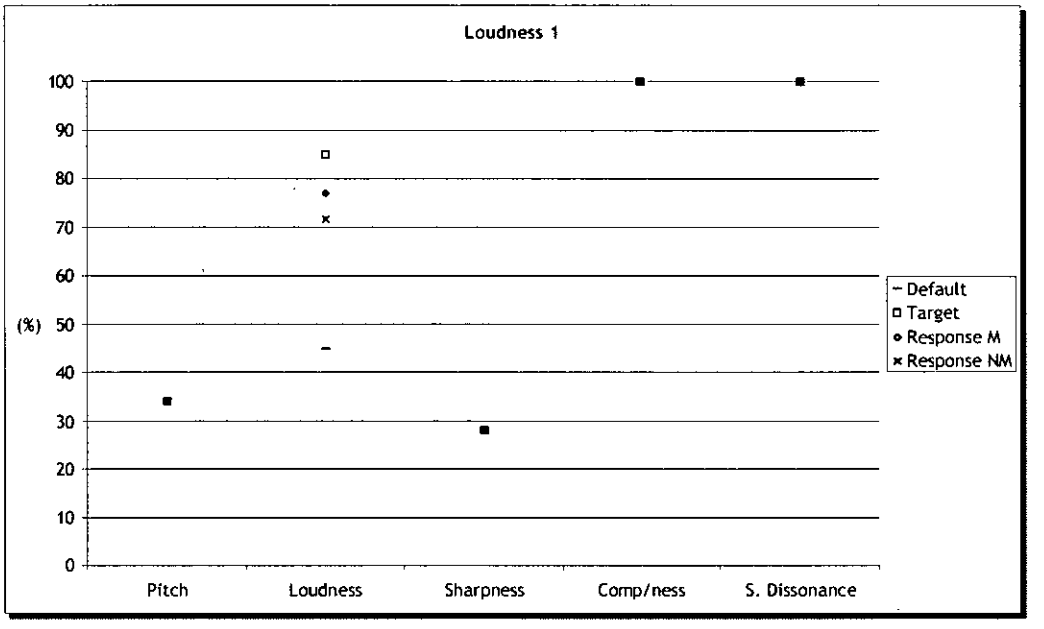
	Strongly Disagree				Strongly Agree
• I think I would like to use *Sound Mosaics* frequently.	1	2	3	4	5
• I found *Sound Mosaics* unnecessarily complex.	1	2	3	4	5
• I thought *Sound Mosaics* was easy to use.	1	2	3	4	5
• I think that I would need the support of a technical person to be able to use *Sound Mosaics*.	1	2	3	4	5
• I found the various functions in *Sound Mosaics* were well integrated.	1	2	3	4	5
• I thought there was too much inconsistency in *Sound Mosaics*.	1	2	3	4	5
• I would imagine that most people would learn to use *Sound Mosaics* very quickly.	1	2	3	4	5
• I found *Sound Mosaics* very cumbersome to use.	1	2	3	4	5
• I felt very confident using *Sound Mosaics*.	1	2	3	4	5
• I need to learn a lot of things before I could get going with *Sound Mosaics*.	1	2	3	4	5

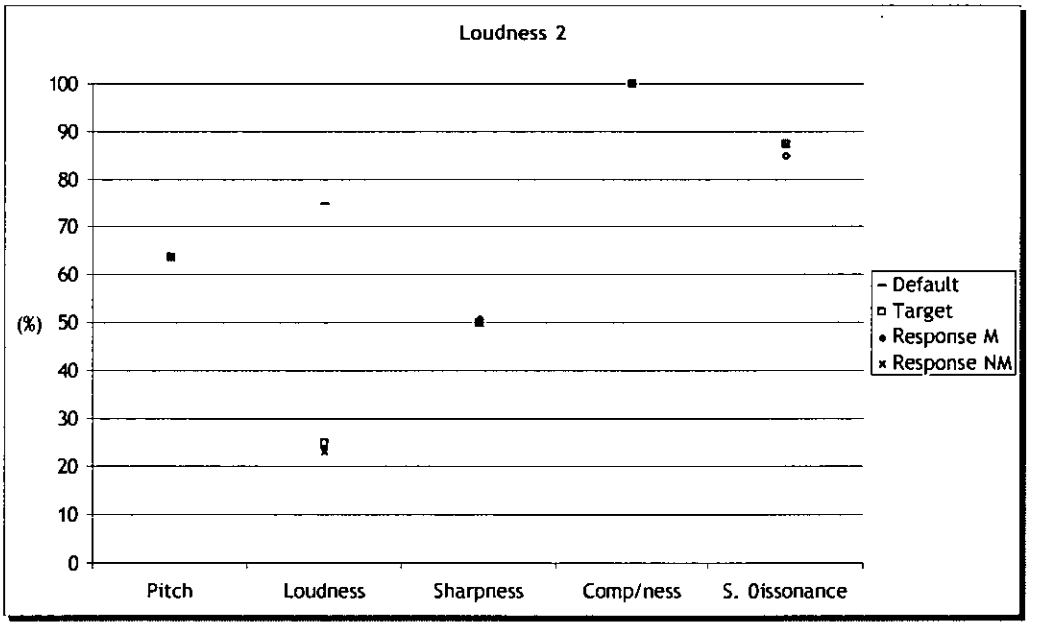
THANK YOU

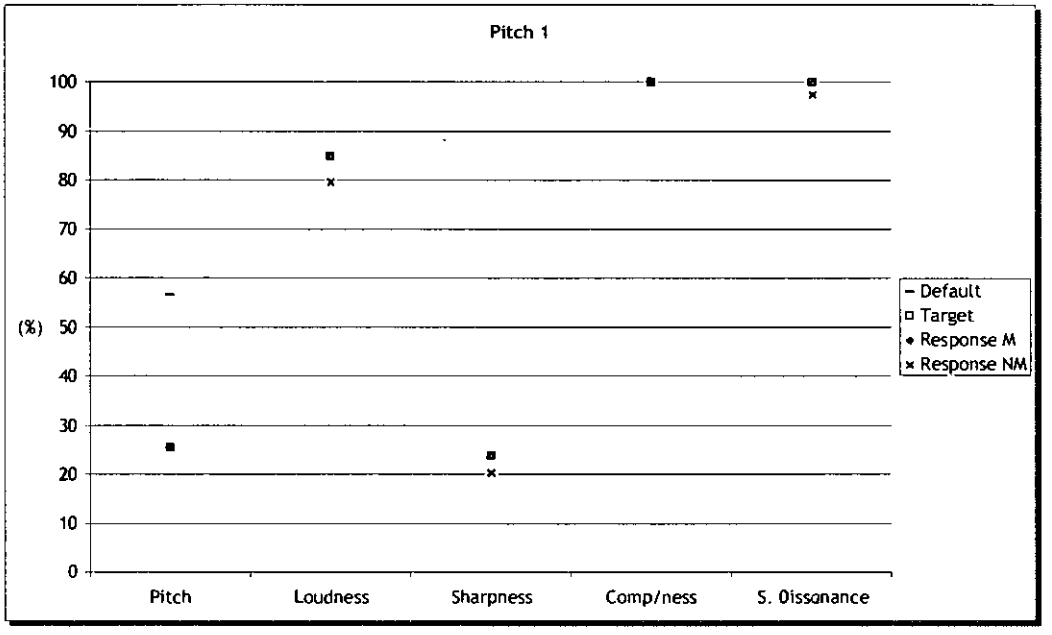
Appendix

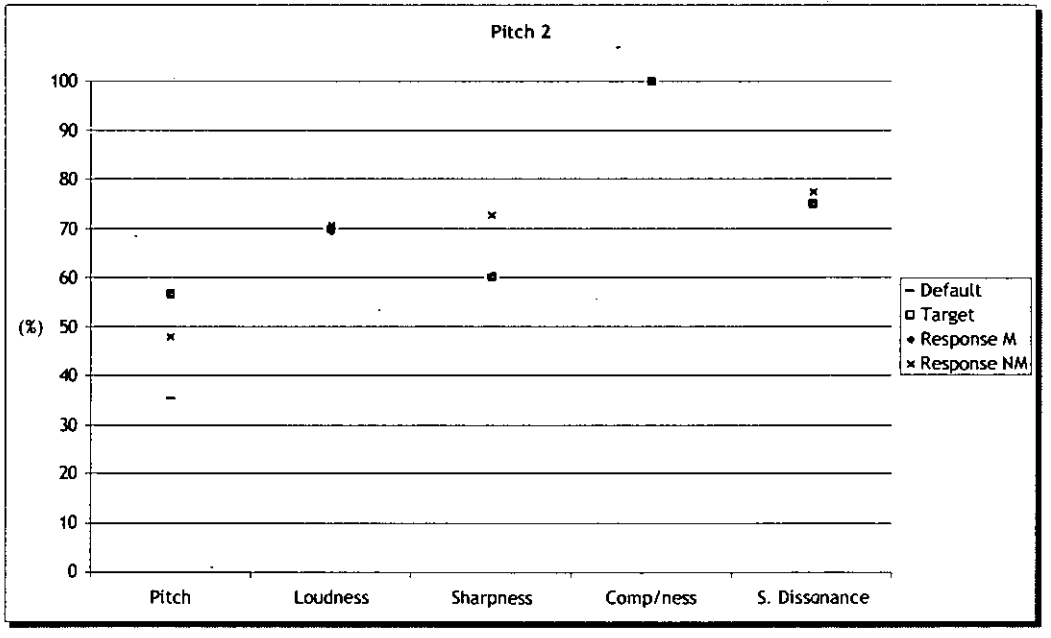
B

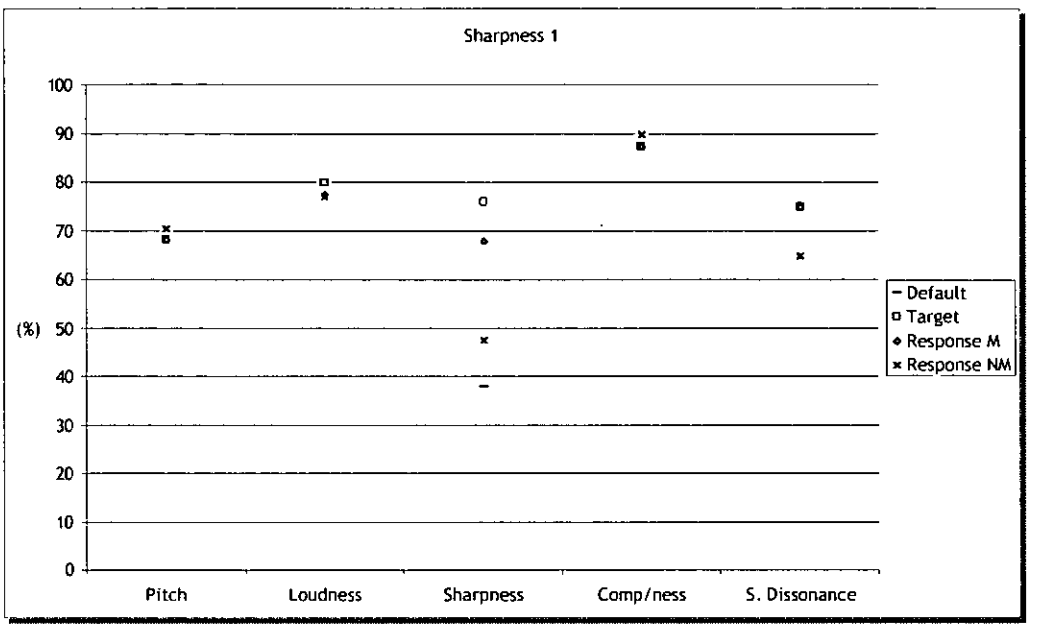
The following figures show accuracy levels for all auditory dimensions in each of the 19 stimuli used in our first usability evaluation described in §7.2.

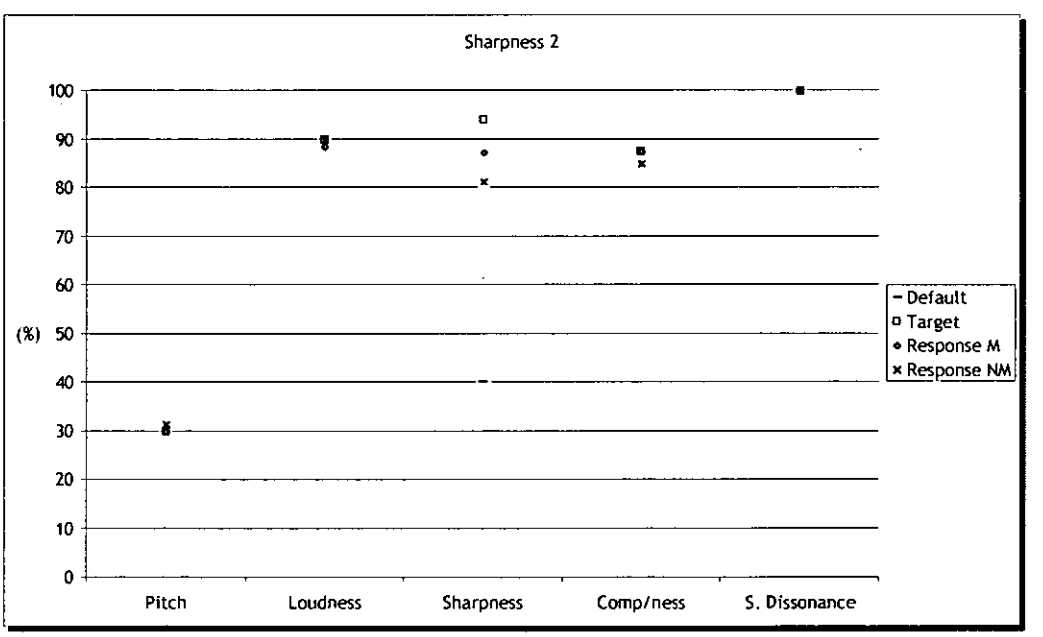


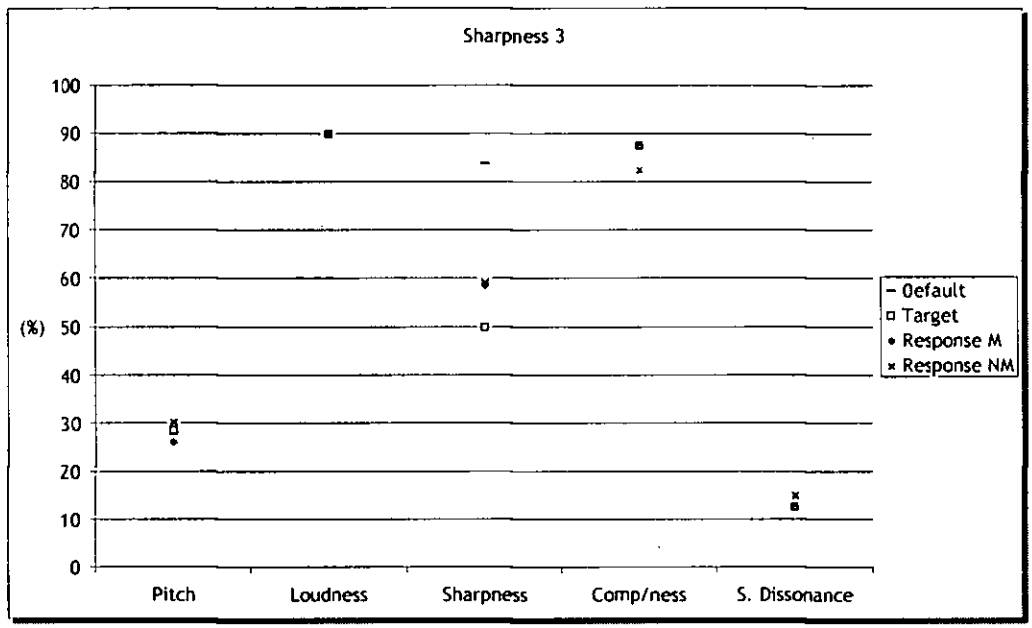


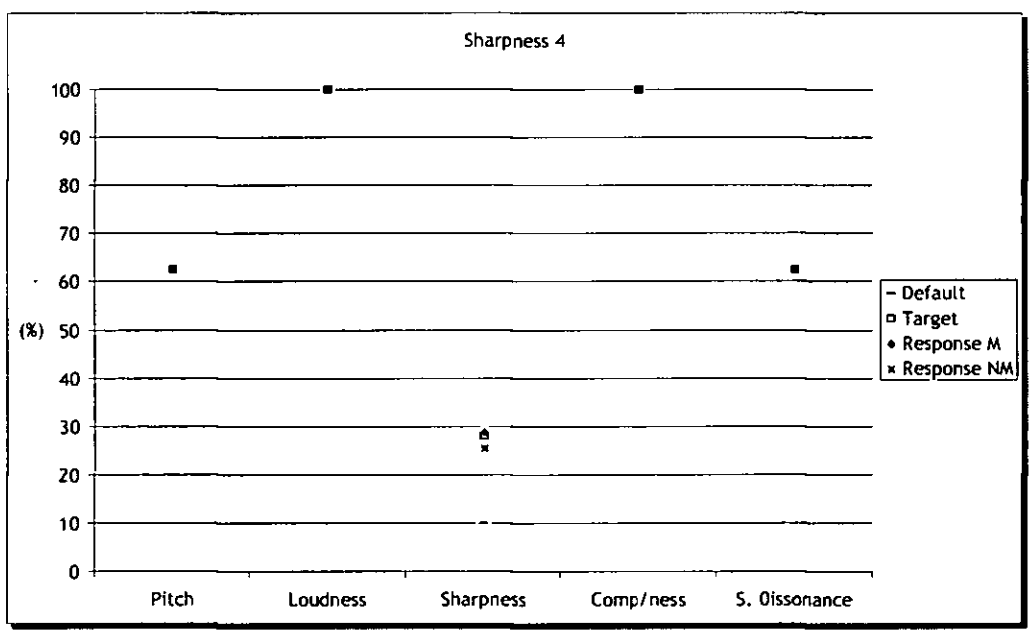

Pitch 1
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Oissonance
Default
Target
Response M
Response NM

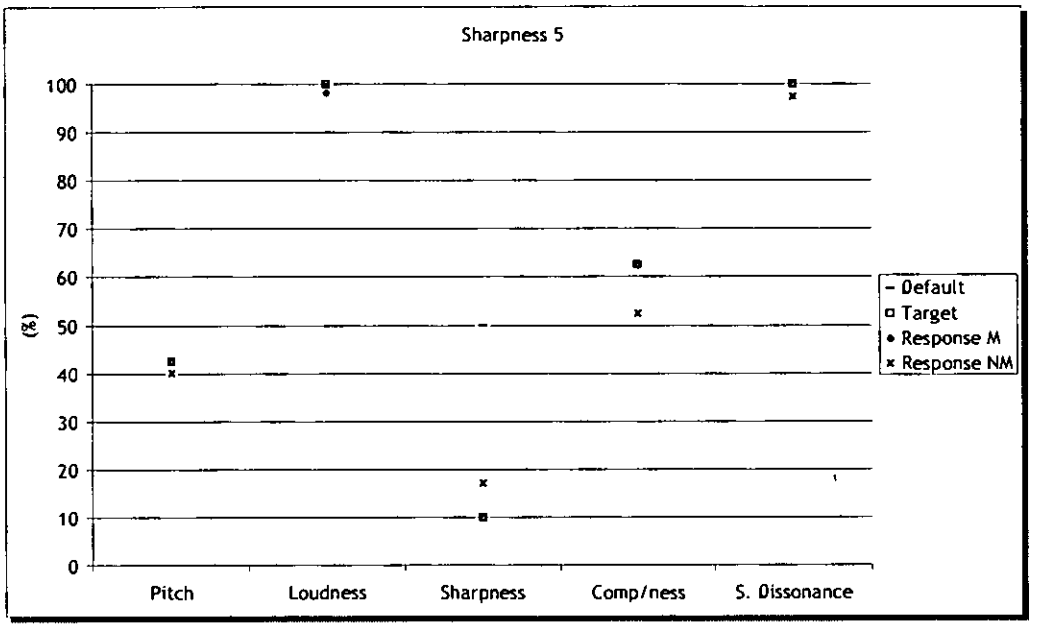

Pitch 2
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

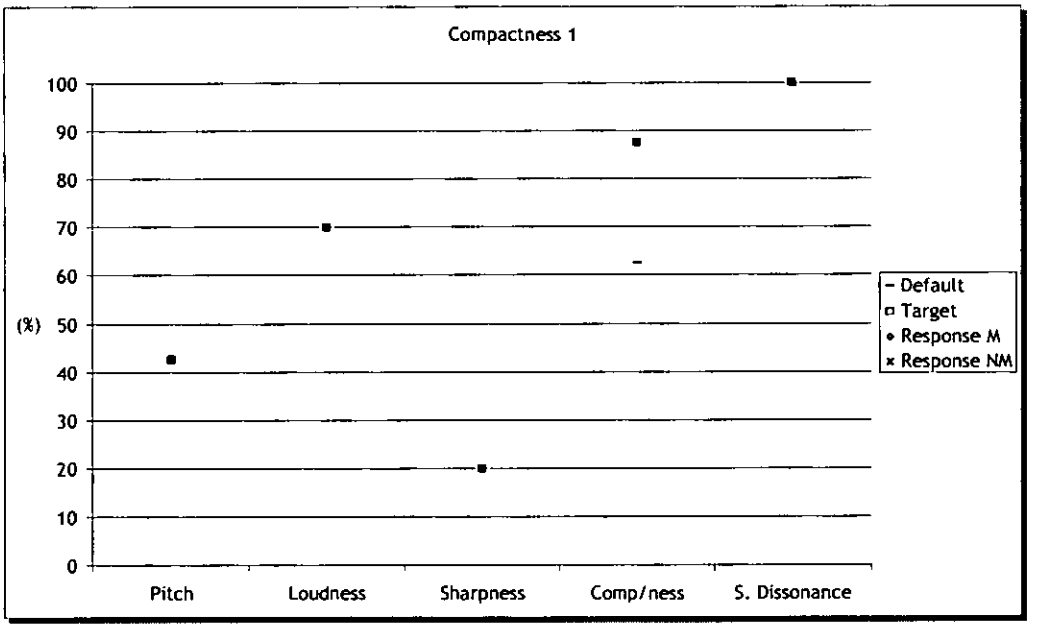

Sharpness 1
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

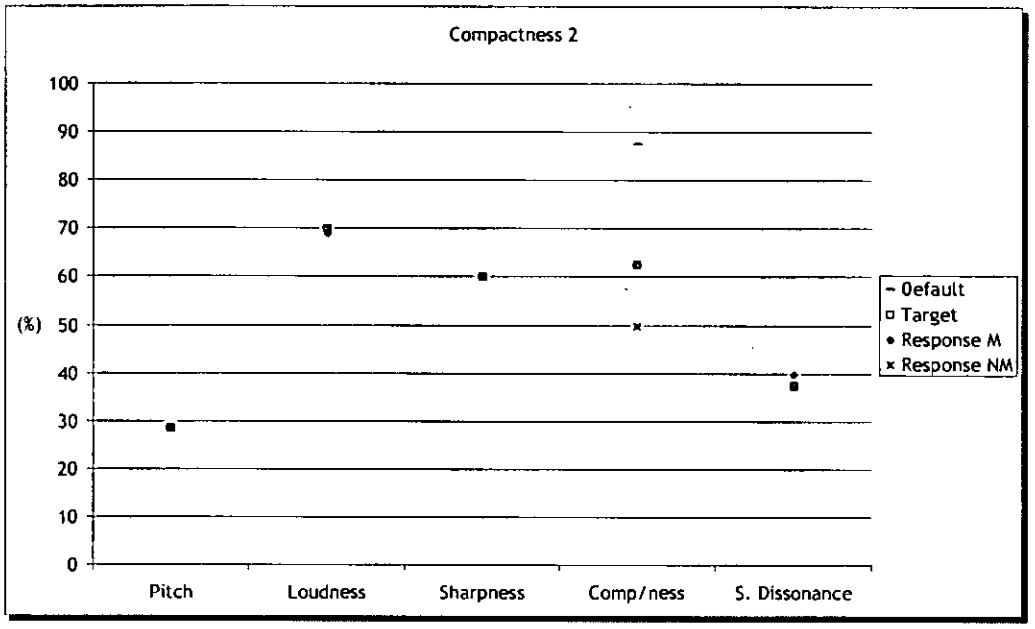

Sharpness 2
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

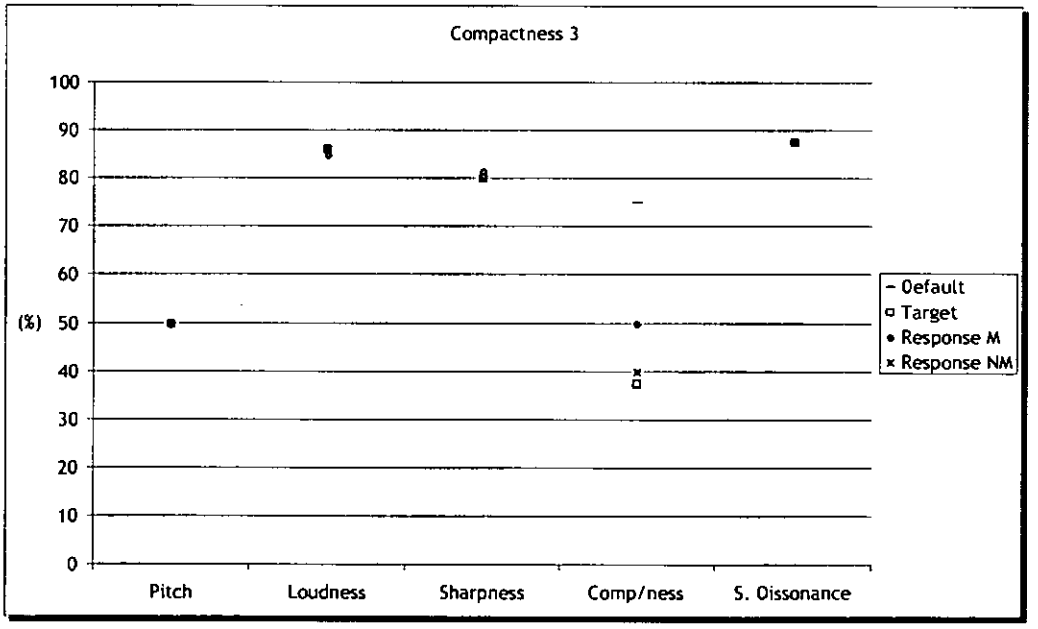

Sharpness 3
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

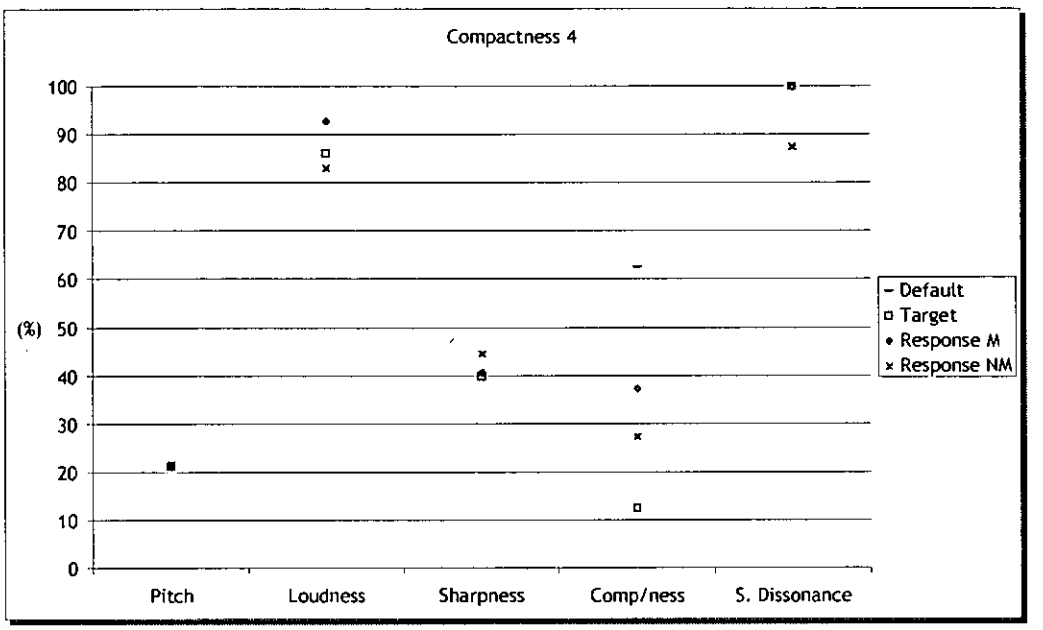

Sharpness 4
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

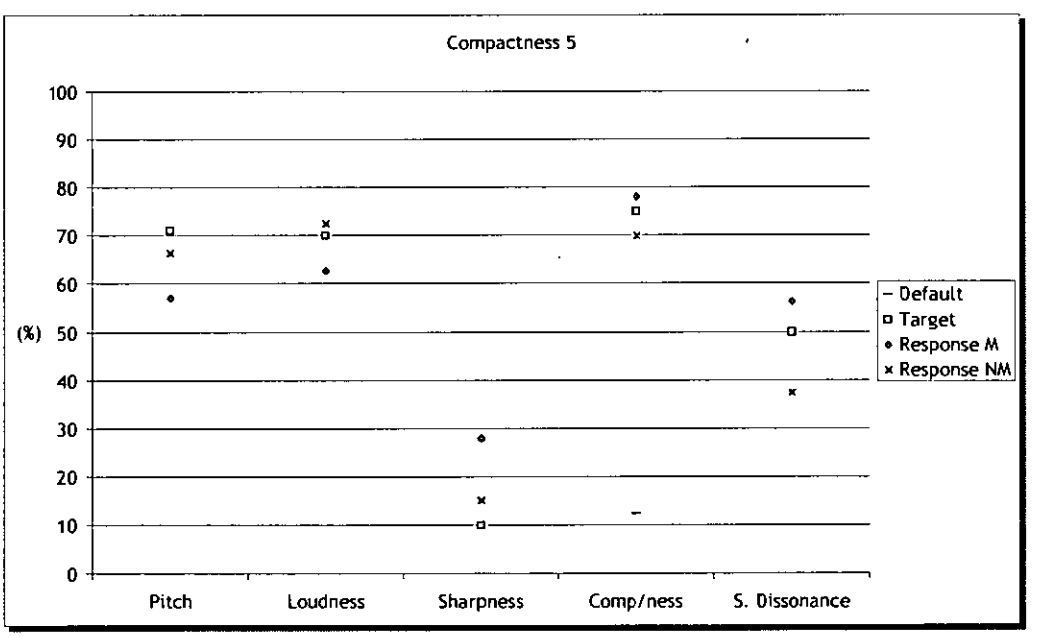

Sharpness 5
100
90
80
70
60
50
40
30
20
10
0
(%)
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

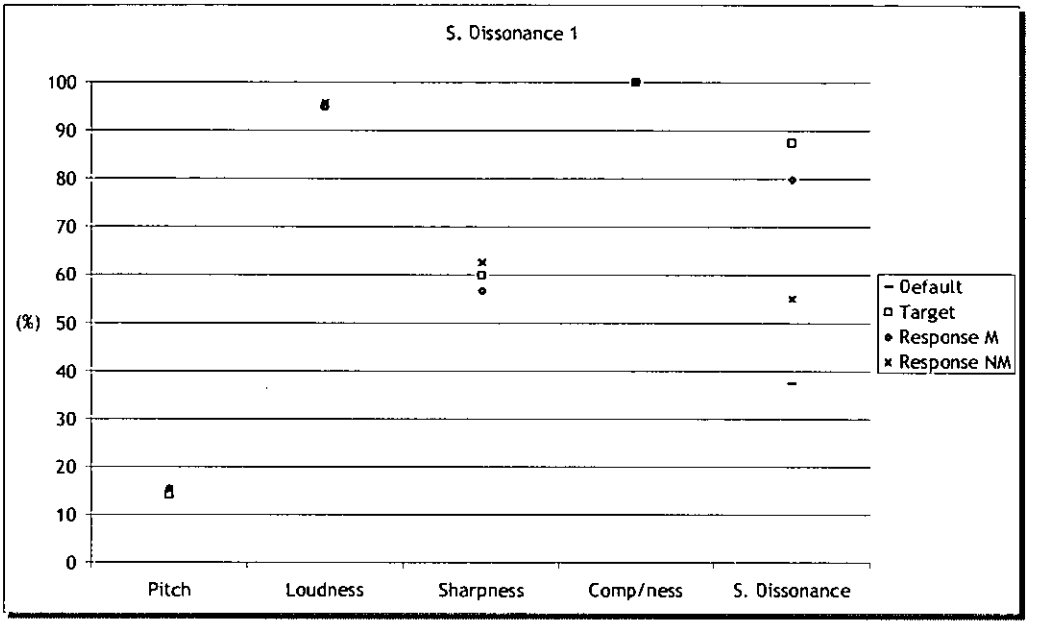

Compactness 1
100
90
80
70
60
50
40
30
20
10
0
(%)
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

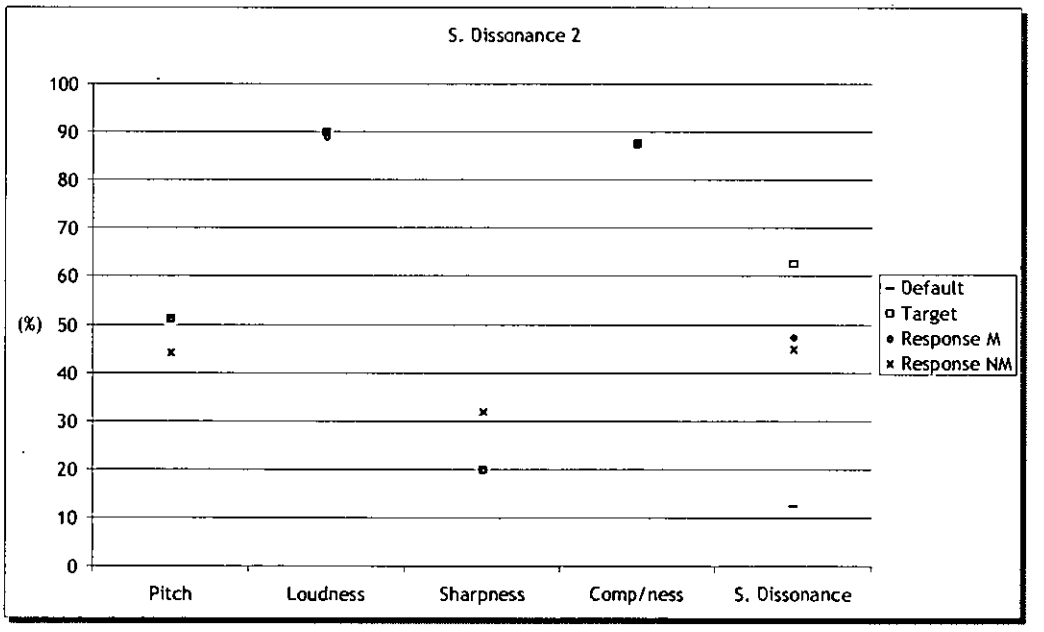

Compactness 2
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

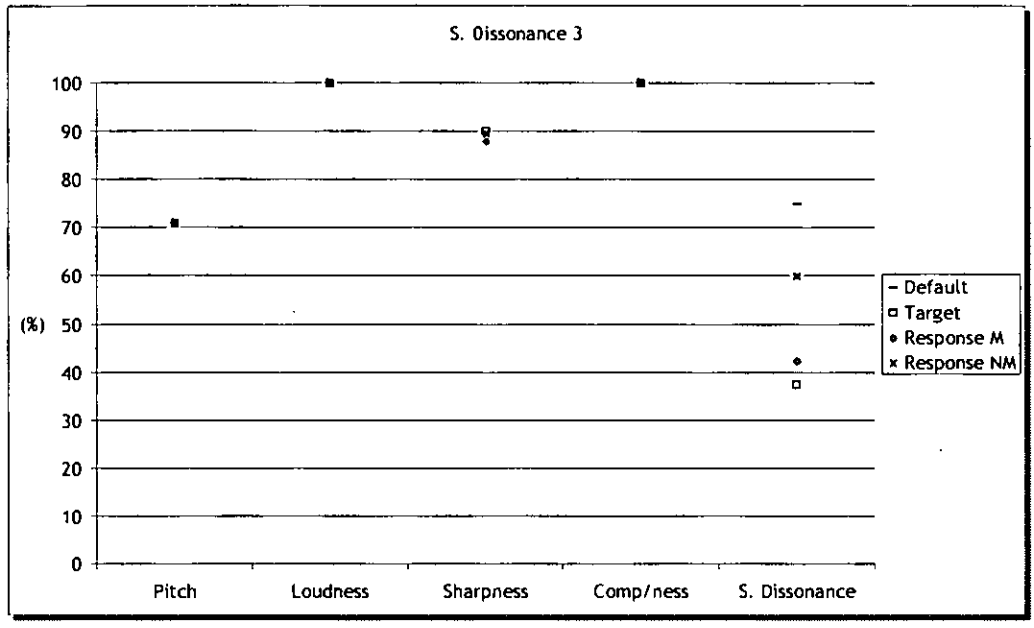

Compactness 3
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

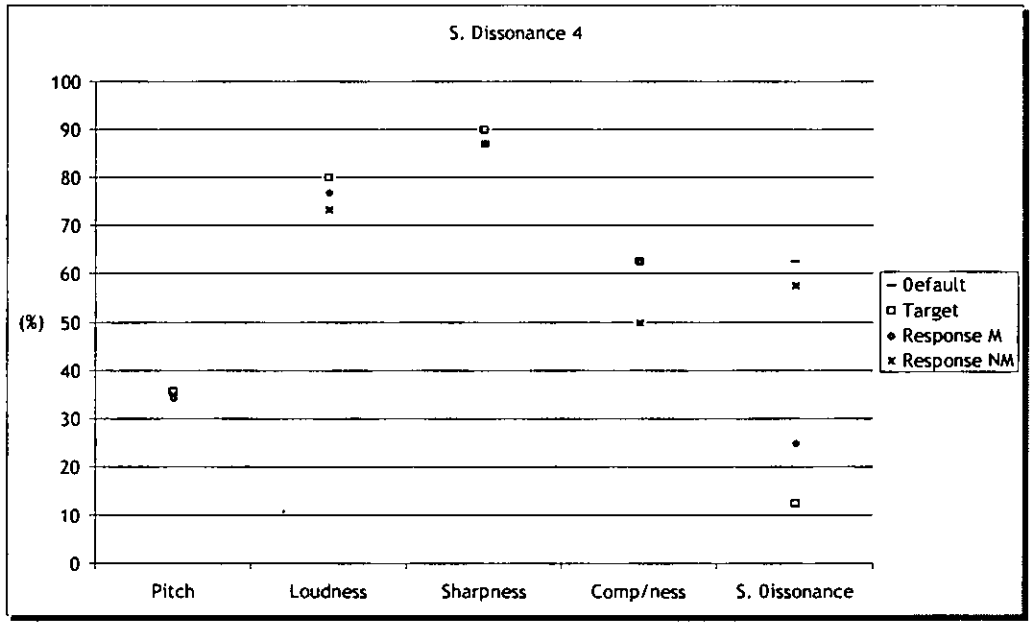

Compactness 4
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

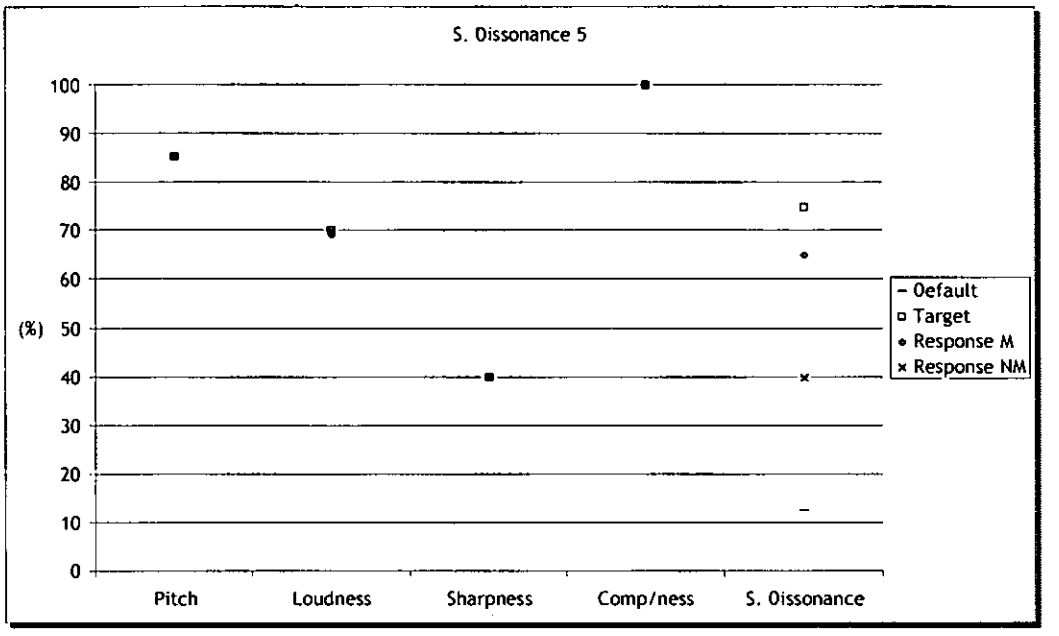

Compactness 5
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM


S. Dissonance 1
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM


S. Dissonance 2
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM


S. Oissonance 3
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

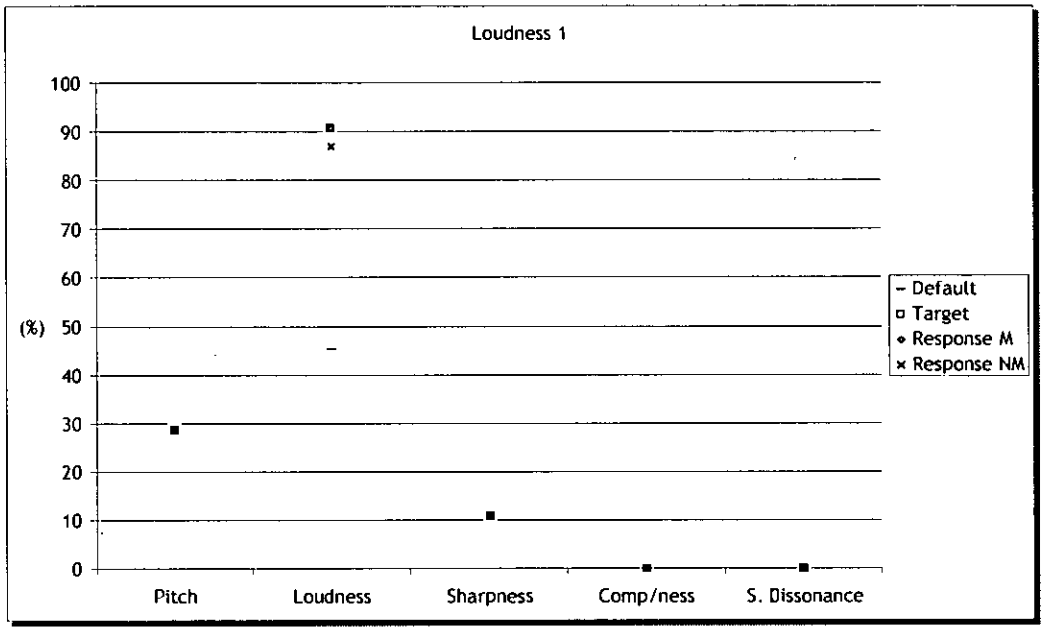

S. Dissonance 4
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Oissonance
Default
Target
Response M
Response NM

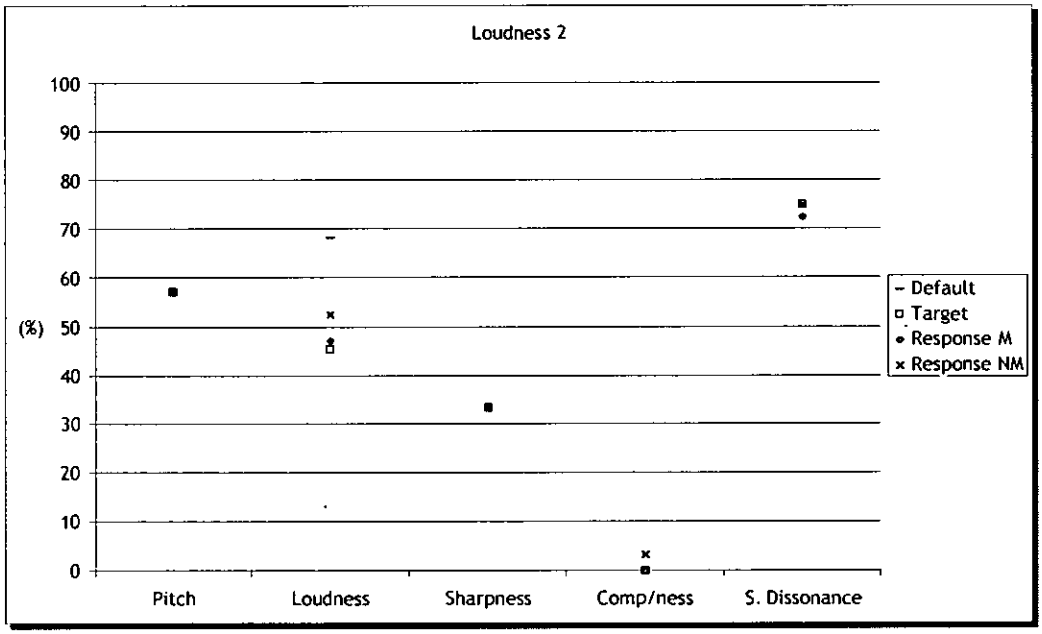

S. Oissonance 5
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Oissonance
Oefault
Target
Response M
Response NM

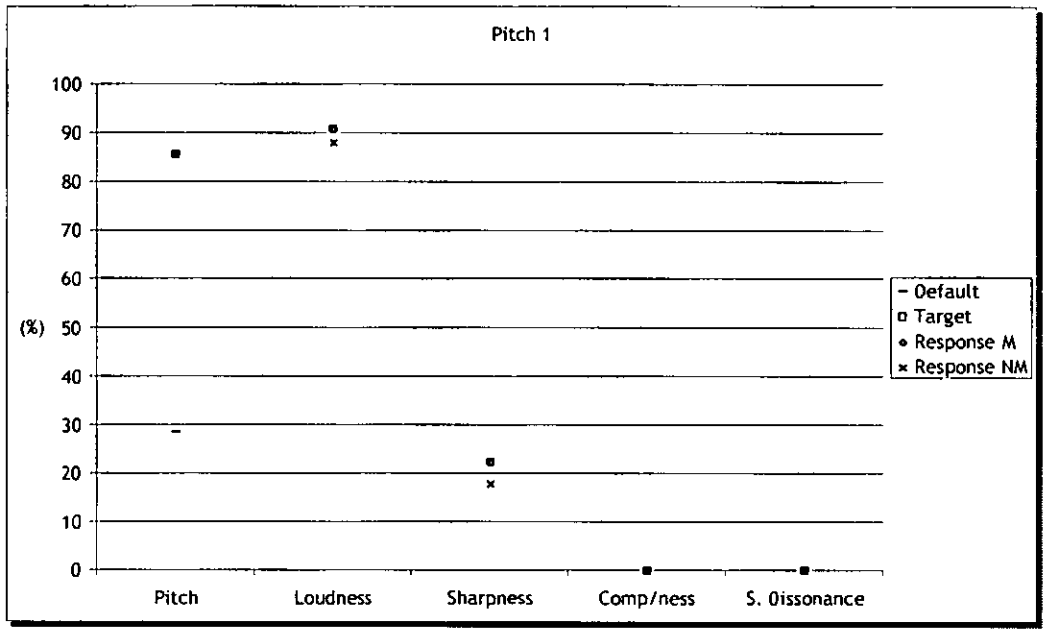
Appendix

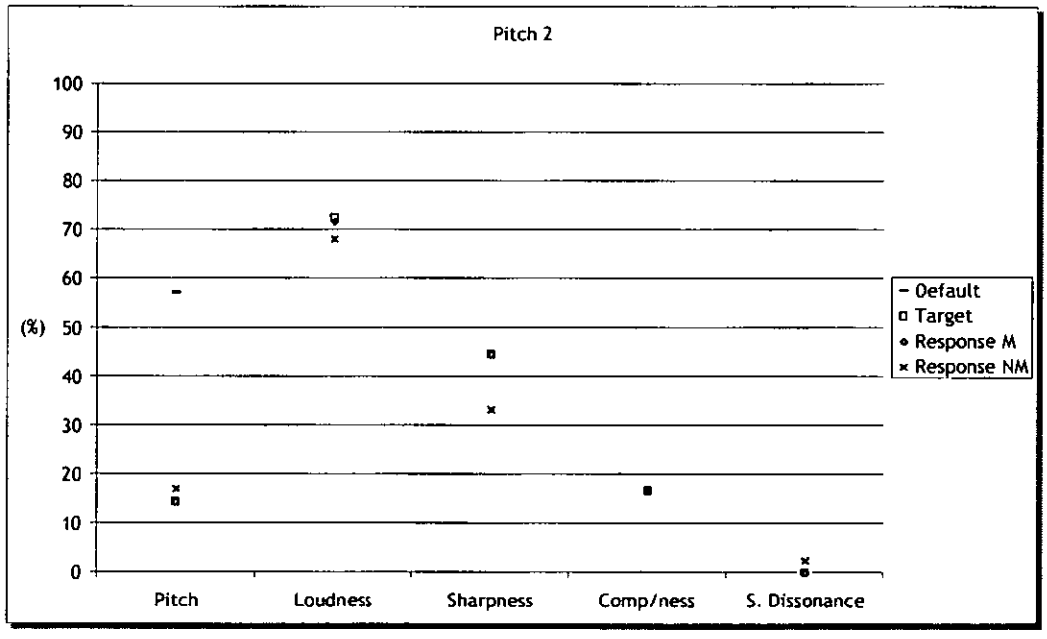
C

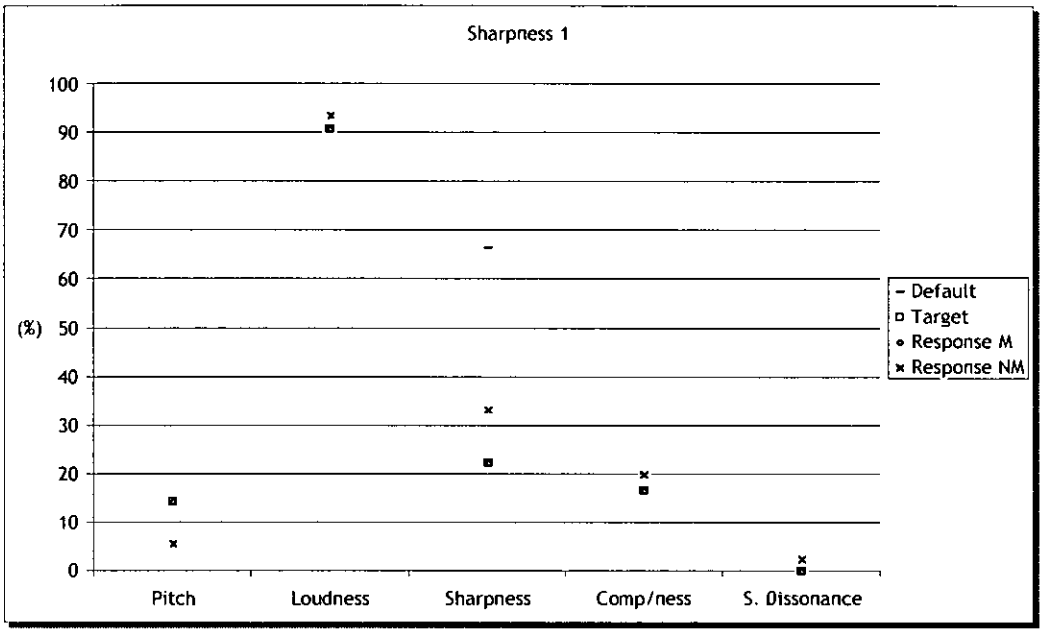
The following figures show accuracy levels for all auditory dimensions in each of the 19 stimuli used in our second usability evaluation described in §8.2.2.

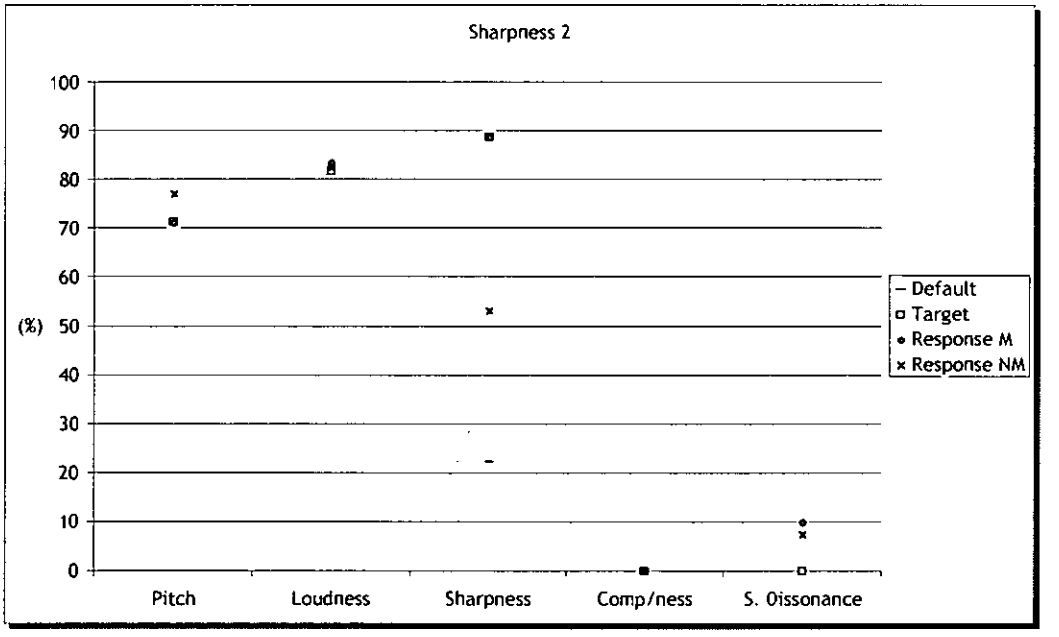


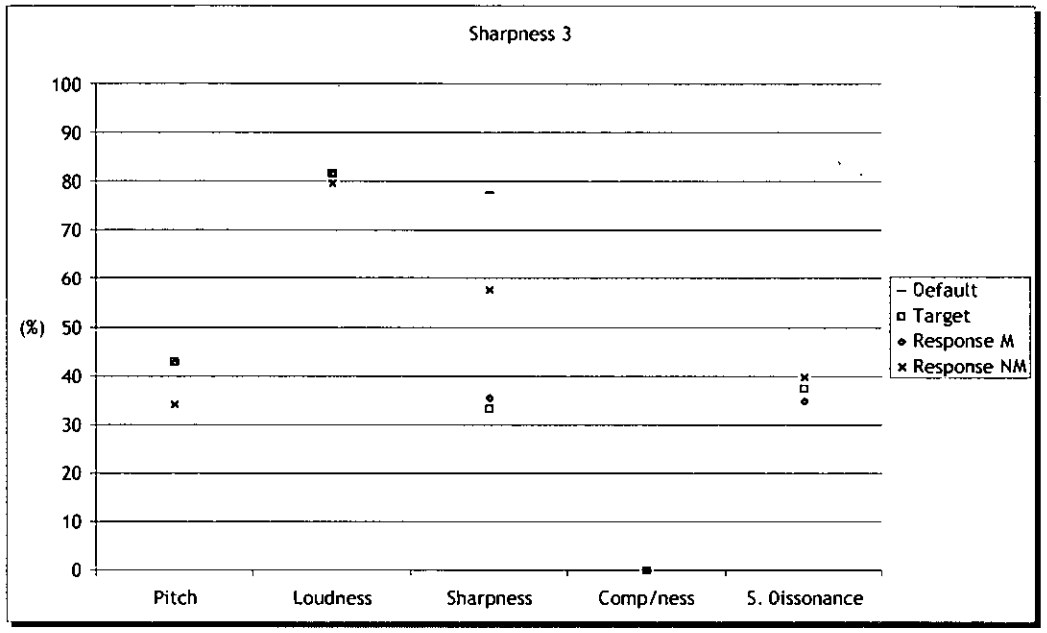


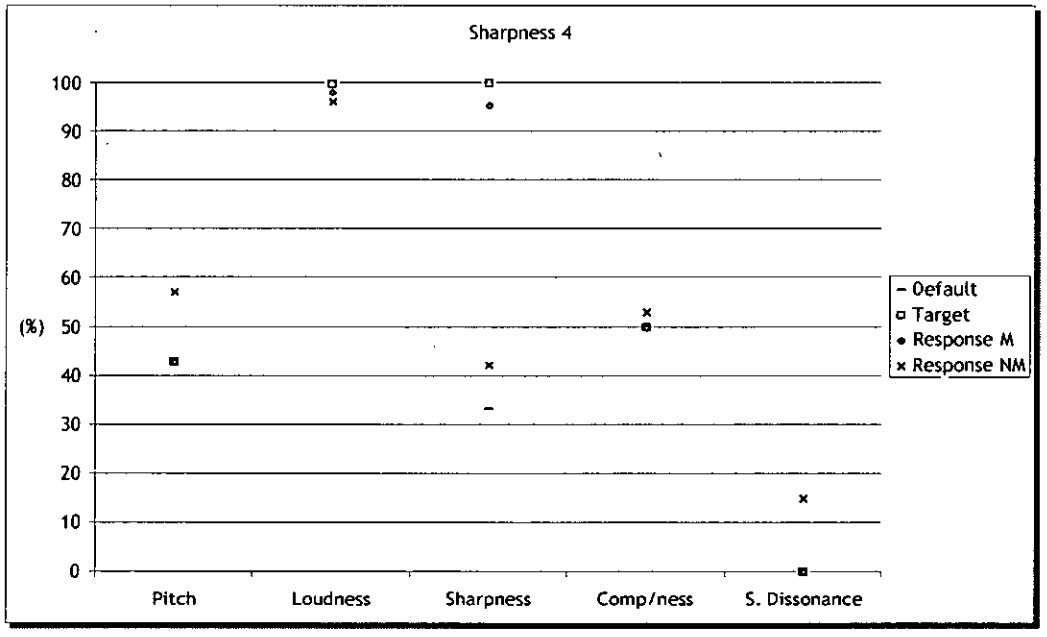

Pitch 1
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Oissonance
- Oefault
Target
Response M
Response NM

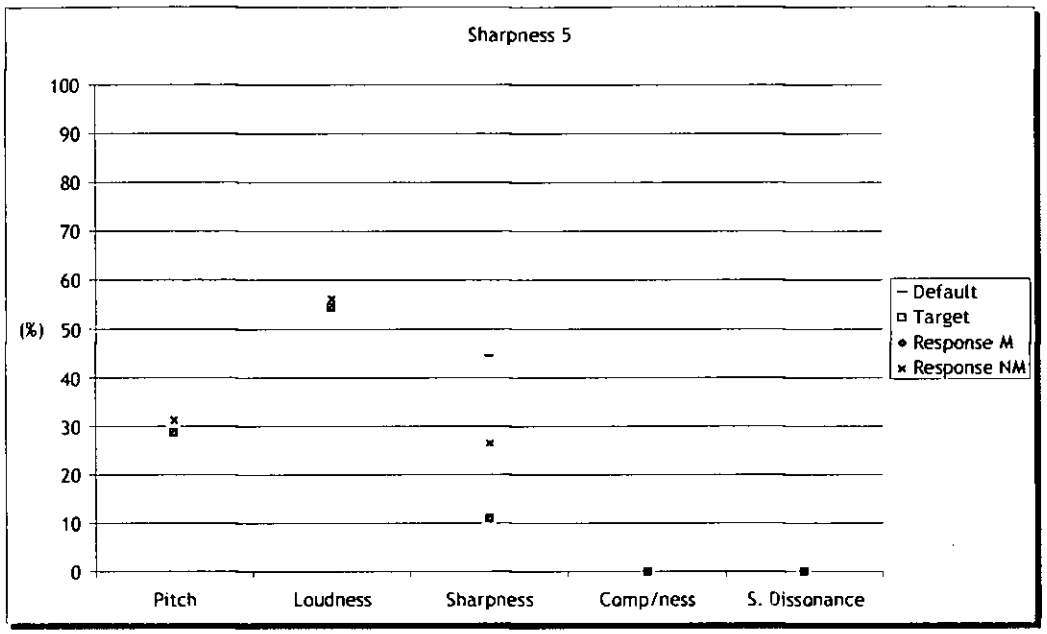

Pitch 2
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
- Oefault
Target
Response M
Response NM

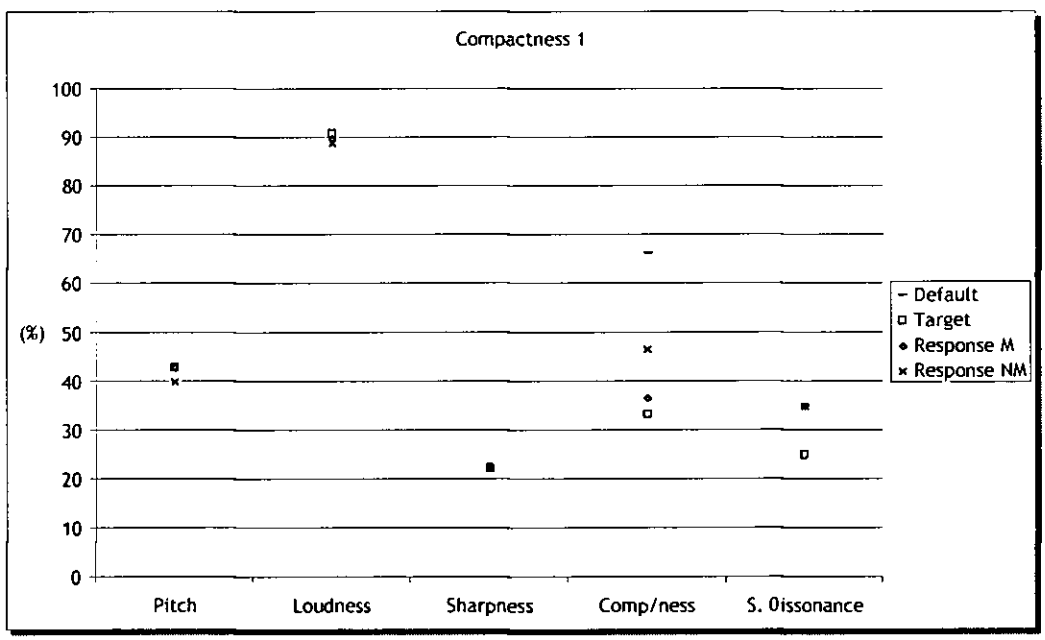

Sharpness 1
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Oissonance
Default
Target
Response M
Response NM

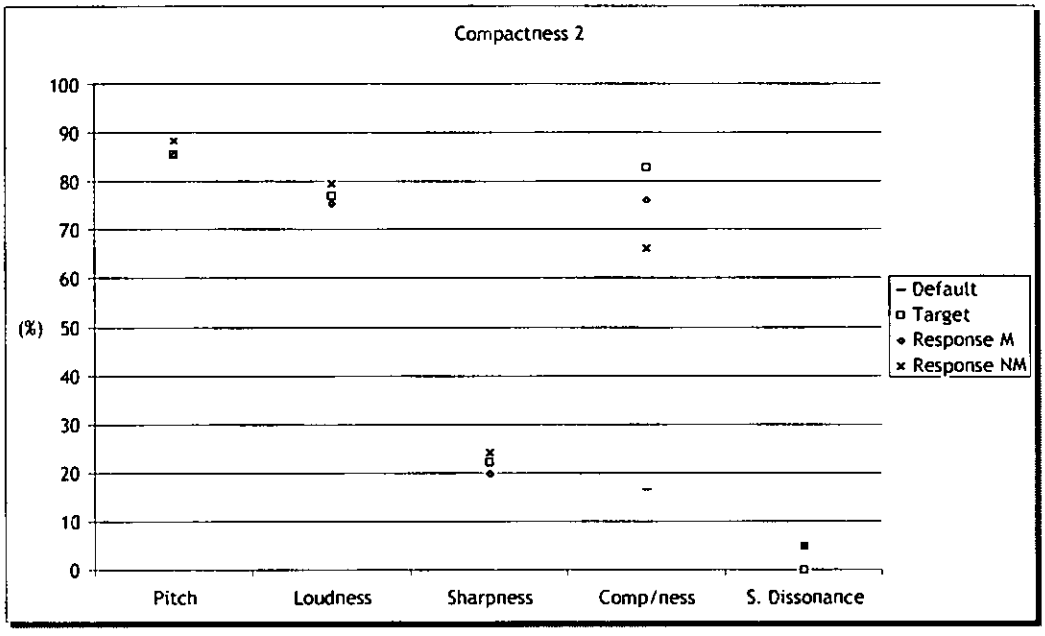

Sharpness 2
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Oissonance
Default
Target
Response M
Response NM

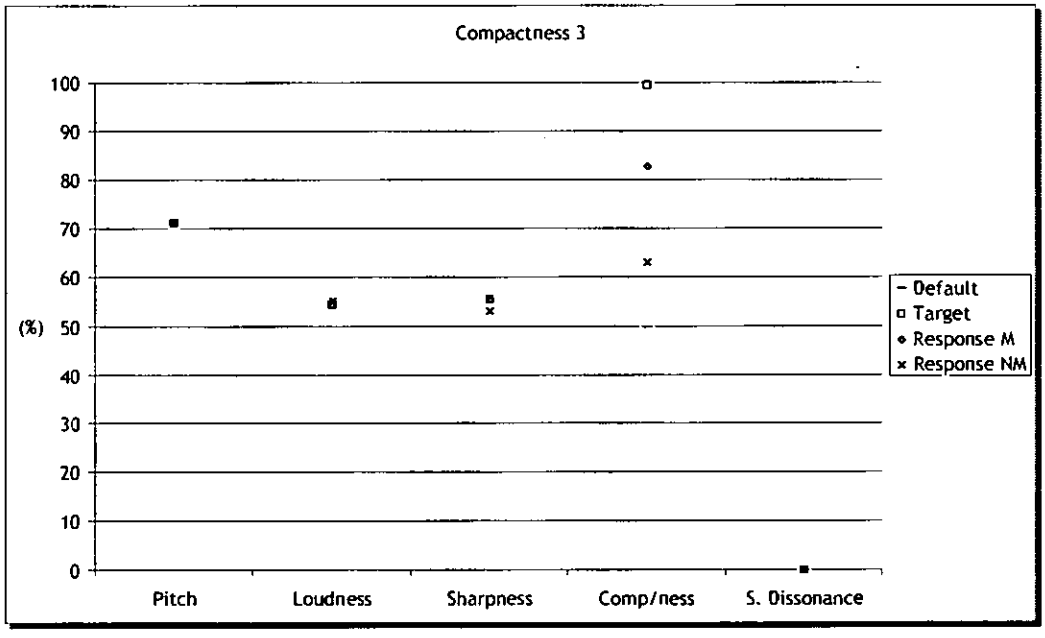

Sharpness 3
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Oissonance
Default
Target
Response M
Response NM

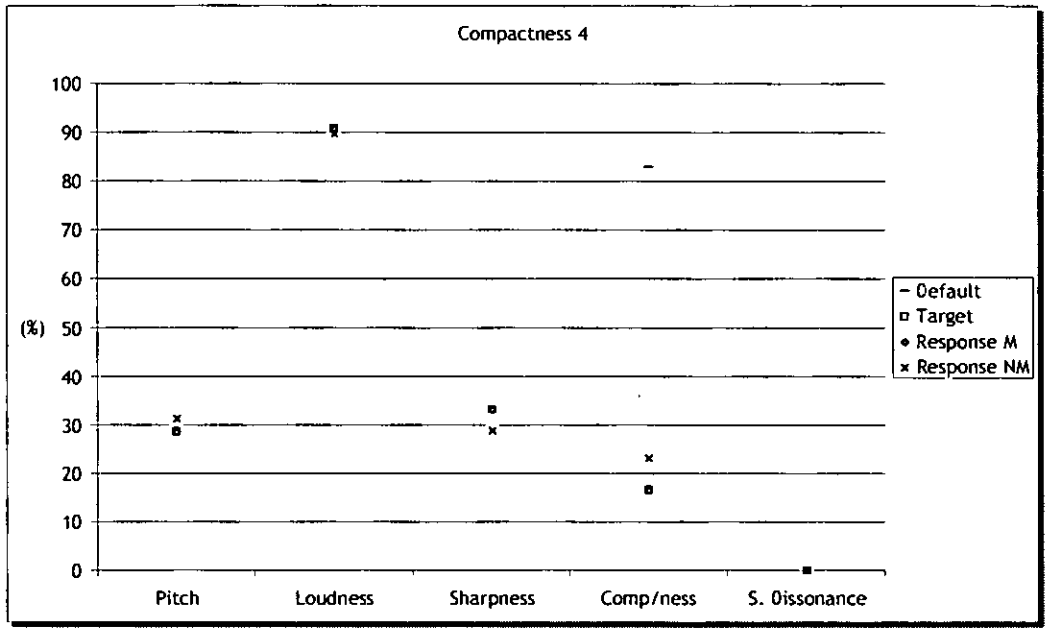

Sharpness 4
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

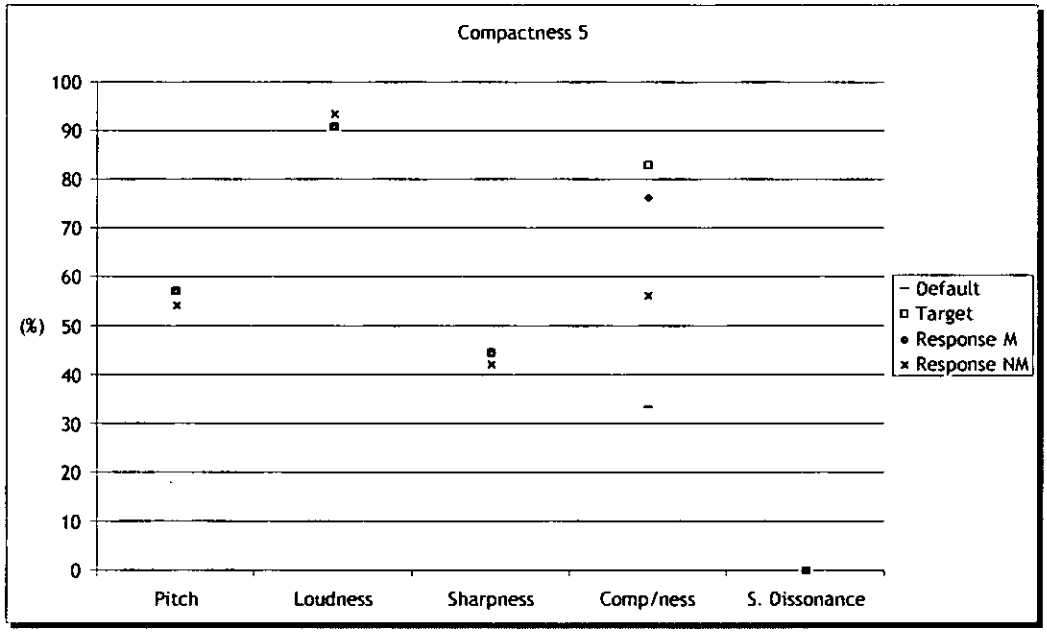

Sharpness 5
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

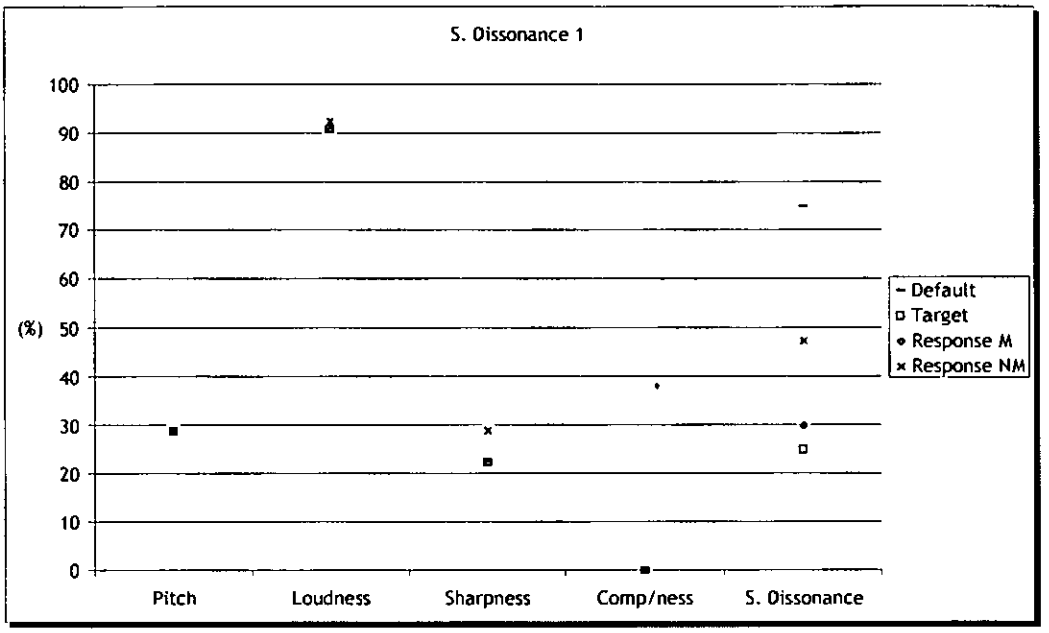

Compactness 1
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

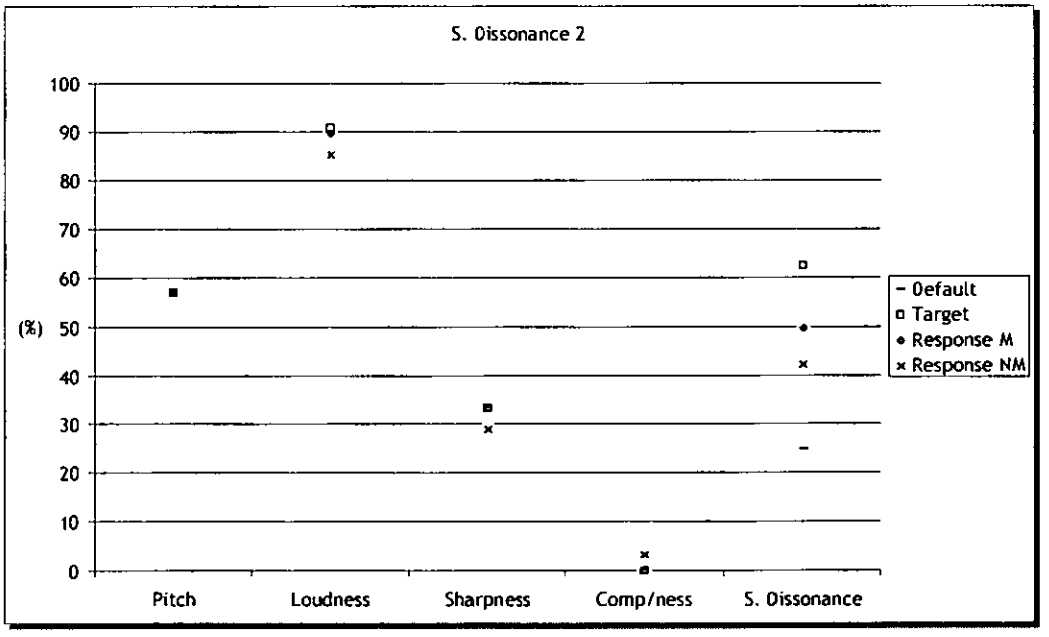

Compactness 2
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

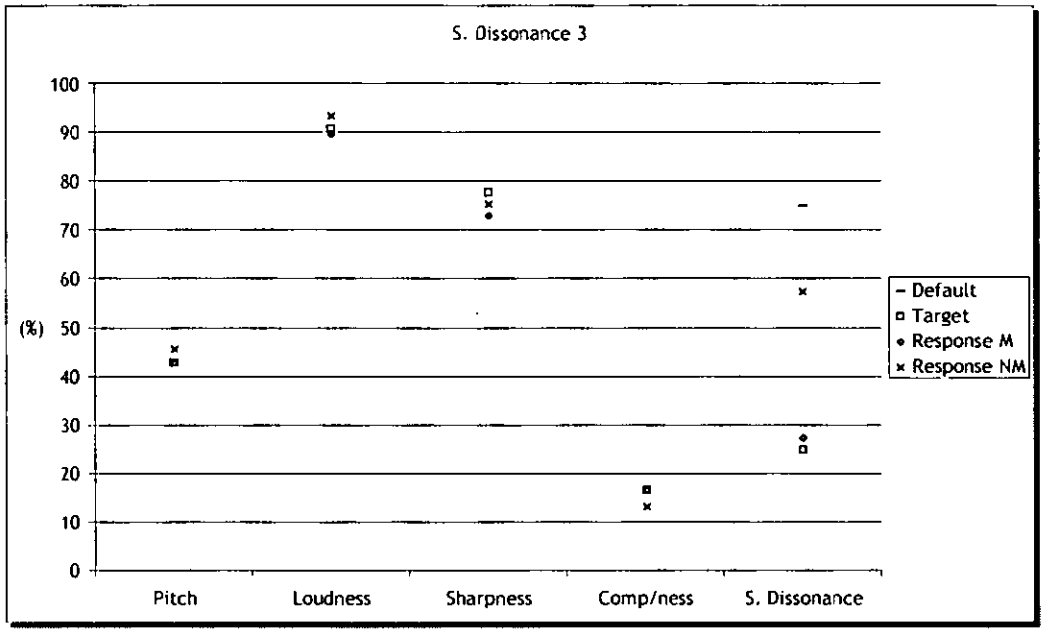

Compactness 3
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

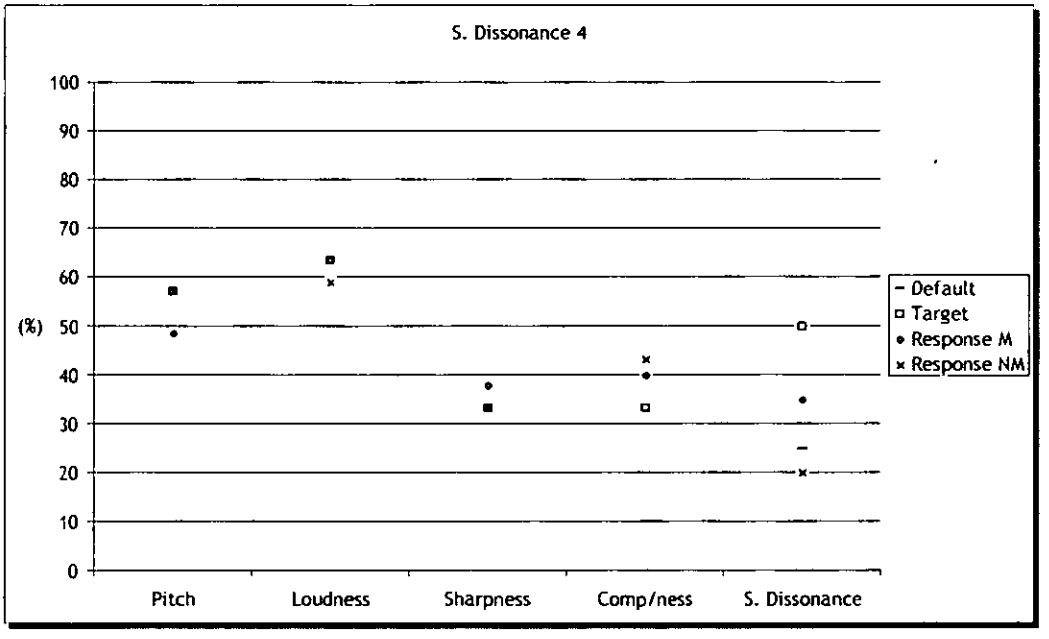

Compactness 4
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM

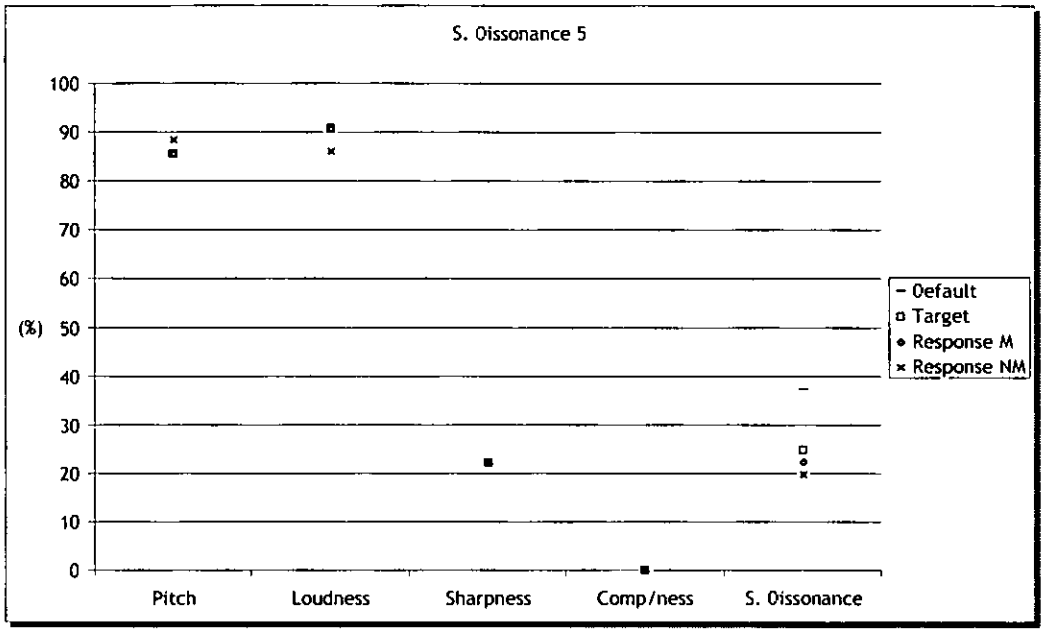

Compactness 5
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM


S. Dissonance 1
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM


S. Dissonance 2
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM


S. Dissonance 3
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM


S. Dissonance 4
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Dissonance
Default
Target
Response M
Response NM


S. Oissonance 5
100
90
80
70
60
(%) 50
40
30
20
10
0
Pitch
Loudness
Sharpness
Comp/ness
S. Oissonance
Oefault
Target
Response M
Response NM

Appendix

D

This is the Csound orchestra file incorporated in the current implementation of Sound Mosaics. Note that the orchestra file is not being modified during the use of Sound Mosaics. When users interact with Sound Mosaics, all sound parameters are written in a Csound score file (see next page) that is constantly being updated to reflect any changes.

Csound Orchestra File

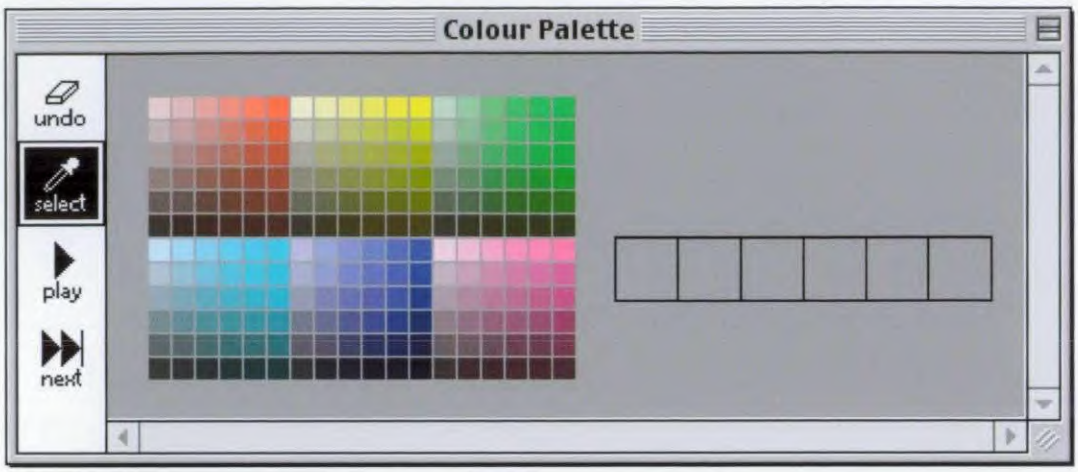
```
sr      = 44100          ; Sampling rate
kr      = 4410           ; Control rate
ksmps   = 10             ; Sampling rate/Control rate Ratio
nchnls  = 2              ; Stereo sound file


; Simple additive synthesis module

        instr 1

kenv  linen  p4,.1,p3,.1     ; Amplitude envelope (reads p4 and p3 from the score file)
a1 oscil kenv, p5, 1         ; Oscillator (reads p5 and function 1 from the score file)

        outs a1,a1           ; Output the result
        endin


; Simple Noise generator module

        instr 2

anoise rand  p4              ; White noise generator (reads p4 from the score file)
afilt reson anoise, p5, p6, 2    ; Bandpass filter (reads p5, p6 from the score file)

        outs afilt, afilt    ; Output the result
        endin
```

This is a sample score file from Sound Mosaics indicating a sound with five partials and five noise bands centred around the partial frequencies (bandwidth = 150 Hz).

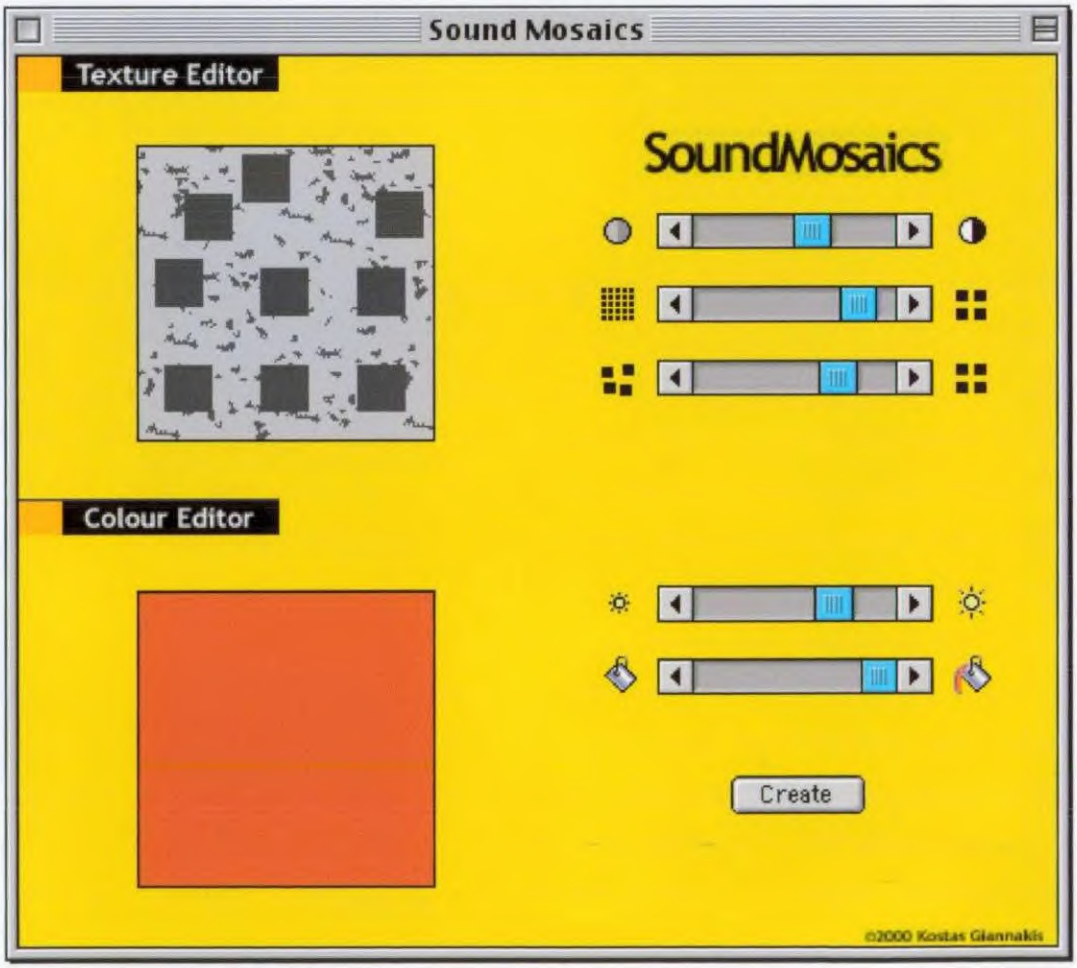
Csound Score File

```
f 1 0 8192 10 1                          ; Function 1

i 1 0 1.5 403 1046.501953125             ; First instance of the additive synthesis instrument
i 2 0 1.5 403 1046.501953125 150         ; First instance of the noise generator centred at the same
                                           frequency as the oscillator above.

i 1 0 1.5 403 2093.00390625              ; Second instance of the additive synthesis instrument
i 2 0 1.5 403 2093.00390625 150          ; Second instance of the noise generator centred at the
                                           same frequency as the oscillator above.

i 1 0 1.5 403 3139.505859375             ; Third instance of the additive synthesis instrument
i 2 0 1.5 403 3139.505859375 150         ; Third instance of the noise generator centred at the
                                           same frequency as the oscillator above.

i 1 0 1.5 403 4186.0078125               ; Fourth instance of the additive synthesis instrument
i 2 0 1.5 403 4186.0078125 150           ; Fourth instance of the noise generator centred at the
                                           same frequency as the oscillator above.

i 1 0 1.5 403 5232.509765625             ; Fifth instance of the additive synthesis instrument
i 2 0 1.5 403 5232.509765625 150         ; Fifth instance of the noise generator centred at the same
                                           frequency as the oscillator above.
```

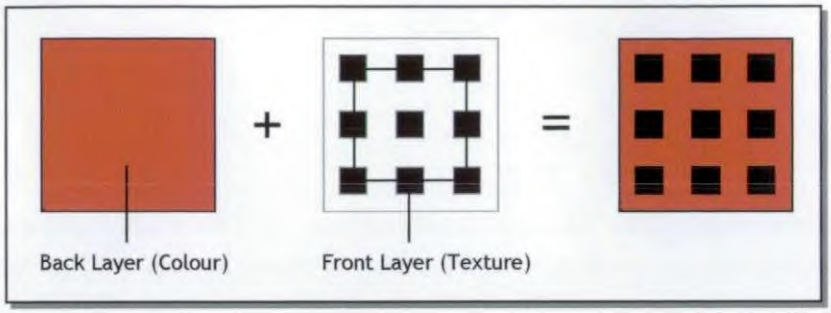
Appendix E



Colour Plate E.1: The prototype application used in the colour-sound experiment described in Chapter 4.

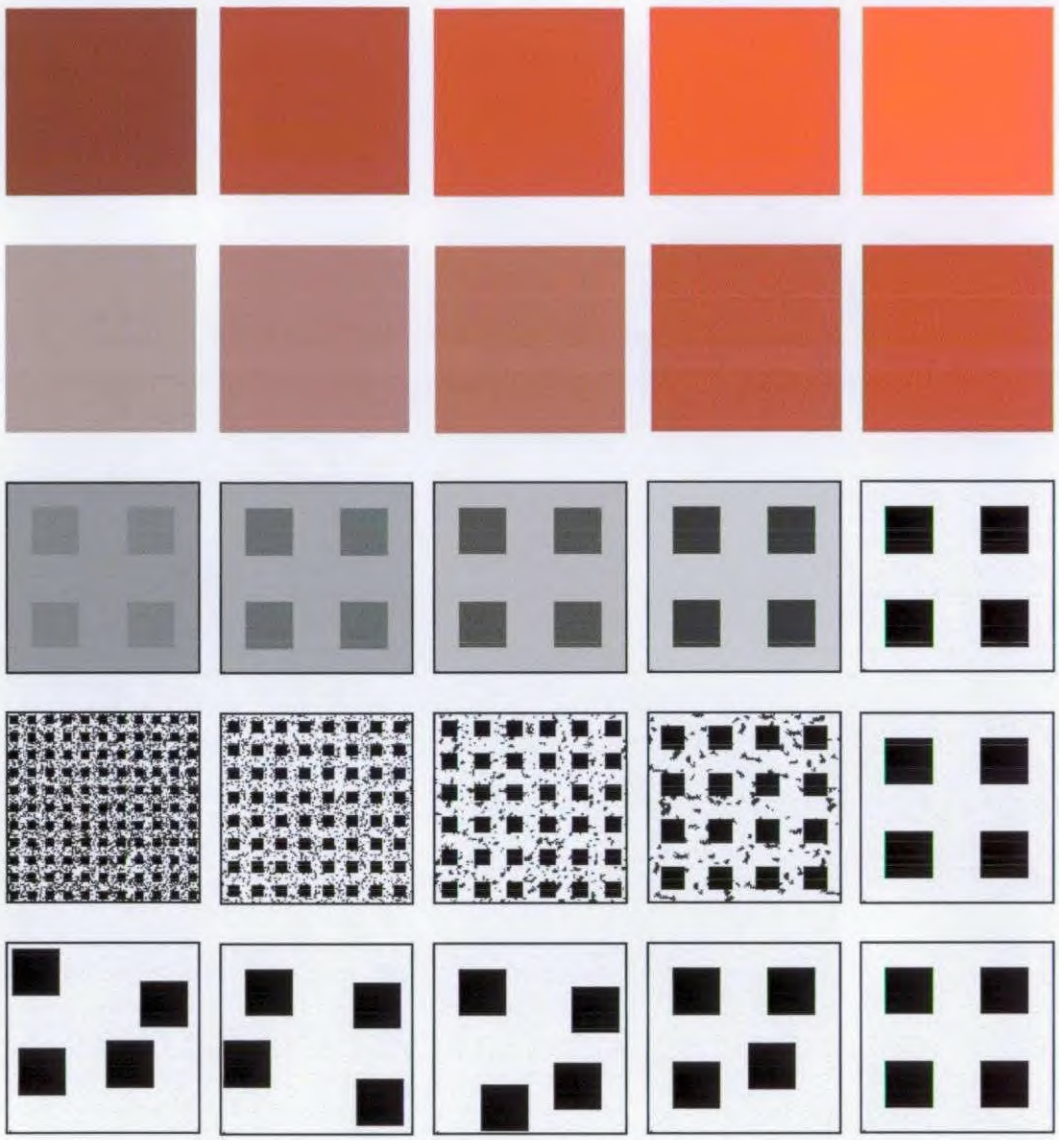


Colour Plate E.2: The initial Sound Mosaics prototype.



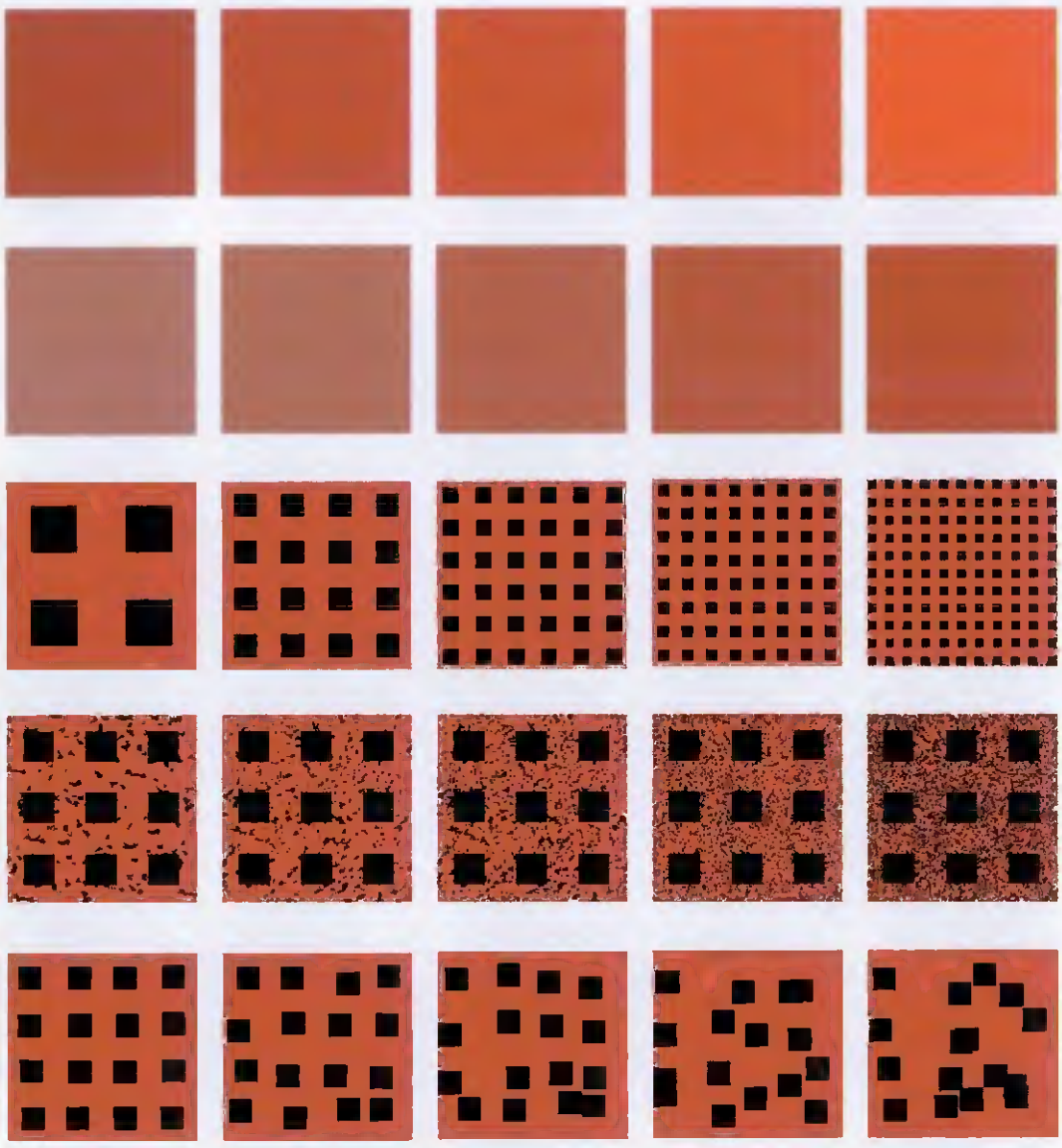
Colour Plate E.3: The figure shows how the colour and texture images were combined in the revised version of Sound Mosaics.

Colour Plate E.4: The revised Sound Mosaics prototype.



Colour Plate E.5: The above visual stimuli were used in the evaluation of the initial implementation of Sound Mosaics to form the content of the image palette when subjects were presented with the Sound Mosaics visualisation framework. From top to bottom: Brightness, Saturation, Contrast, Coarseness/Granularity, and Periodicity.

Colour Plate E.6: The above visual stimuli were used in the evaluation of the initial implementation of Sound Mosaics to form the content of the image palette when subjects were presented with the frequency-domain visualisation framework. From top to bottom: Height, Brightness, Line addition, Pixelation, and Density.



Colour Plate E.7: The above visual stimuli were used in the evaluation of the revised implementation of Sound Mosaics to form the content of the image palette when subjects were presented with the Sound Mosaics visualisation framework. From top to bottom: Brightness, Saturation, Coarseness, Granularity, and Periodicity.

Colour Plate E.8: The above visual stimuli were used in the evaluation of the revised implementation of Sound Mosaics to form the content of the image palette when subjects were presented with the frequency-domain visualisation framework. From top to bottom: Height, Brightness, Line addition, Pixelation, and Density.